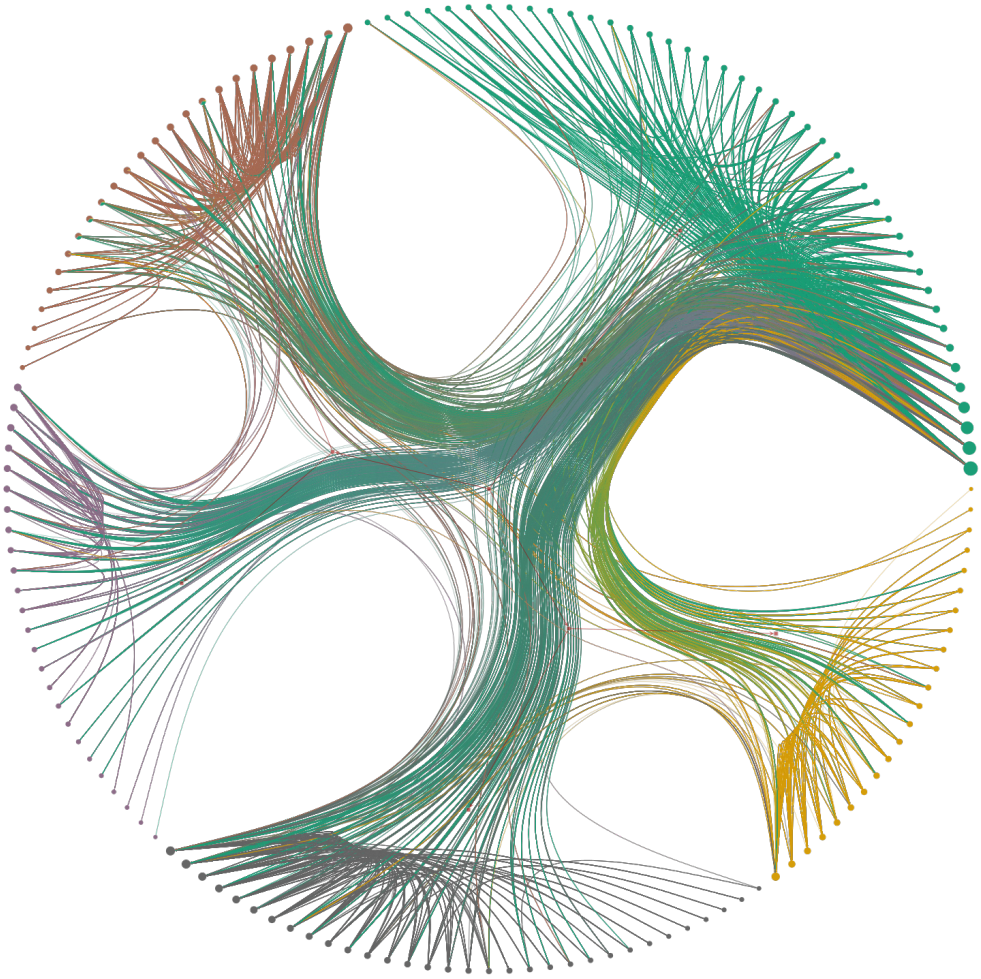
FACULTY OF SCIENCES

FACULTY OF ECONOMICS AND BUSINESS ADMINISTRATION

A temporal network perspective of collective behavior in economic systems



GHENT UNIVERSITY

Benjamin Vandermarliere

FACULTY OF SCIENCES

FACULTY OF ECONOMICS AND BUSINESS ADMINISTRATION

A temporal network perspective of collective behavior in economic systems

Supervisor: Prof. Dr. Jan Ryckebusch
Co-supervisor: Prof. Dr. Koen Schoors

Dissertation submitted in fulfillment
of the requirements for the degree of:
Doctor of Science: physics and applied economic sciences

Ghent University
Faculty of Sciences:
Department of Physics and Astronomy
Faculty of Economics and Business Administration:
Department of General Economics
Academic year 2016-2017

The figure on the cover depicts the historical trade network of
Chapter 3 during the interbellum.

GHENT UNIVERSITY

Benjamin Vandermarliere

Dankwoord

tl;dr: Thanks everyone for helping me out! :)

De afgelopen vijf jaar had ik het geluk om heel erg goed omringd te zijn. Zonder de hulp van vele mensen was dit doctoraat er nooit geweest.

Als eerste wil ik natuurlijk Jan bedanken. Bedankt om mij als masterstudent onder je vleugels te nemen en om het eerste jaar van mijn doctoraat te financieren. Bedankt voor alle raad en ook om telkens de laatste nodige push te geven aan onze papers. Naast Jan had ik ook de eer om Koen als promoter te hebben. Bedankt om met ons fysici in zee te willen gaan, om steeds weer te inspireren en enthousiasme op te wekken, en voor alle kansen die je me hebt aangereikt.

De man die Jan en Koen heeft samengebracht is natuurlijk ook cruciaal in dit verhaal. Rogier, ik keek altijd erg uit naar je vele visionaire verhalen en de daaropvolgende brainstormsessies.

Gelukkig kon ik bij ieder project rekenen op mijn co-auteurs. Alexei, thanks for kick-starting my first paper with your meticulously cleaned data and the later help with the write-up. Peter, het ETF project en het bijhorende bezoek aan Zwitserland heb ik aan jou te danken. Bedankt voor die mooie ervaring! Didier, thanks for hosting me in Zurich and for your continuous stream of ideas. Simon, het was een plezier om samen met jou en Jan je uitstekende masterthesis in een paper te gieten.

De meest productieve samenwerking was ongetwijfeld die met mijn twee top kantoorgenoten. Stijn, bedankt om ons project op poten te zetten en alles historisch te kaderen. Sam, je technisch vernuft en schrijfskills verbazen mij nog steeds. Merci kerels!

The final co-author who contributed significantly to my PhD is Aaron. Your visits were true heydays of productivity. Thanks for transferring all that knowledge and being a mentor in my first years. I will also never forget my visit to Tokyo. Thanks dude, you're awesome!

Alhoewel hij technisch gezien nog geen co-auteur is, wil ik hier graag toch al Ken bedanken. Onze samenwerking bij de bank was een waar genoegen. Bedankt om dat contract rond te krijgen, voor het delen van je vele business wijsheden en je niet aflatend aanstekelijk enthousiasme. Thanks buddy! Hopelijk gaat je post-it allergie ooit nog weg ...

Toen ik enkele jaren geleden begon als enige 'interdisciplinaire' doctoraatsstudent van Jan en Koen, had ik nooit durven dromen dat er nog veel zouden volgen:

de LAB-Mers. Het was steeds erg fascinerend om tijdens een meeting te praten over tritanium en ruimteschepen :) Thanks Kevin, Milan and Andres to all be a part of the EVE space opera adventure.

Het beste aspect van een interdisciplinair doctoraat is de kans om op twee verschillende plekken te mogen vertoeven. In mijn geval was dit het INW en 'op de economie'.

Camille en Tom, samen hebben we heel de rit doorlopen. Bedankt om geregeld naar mijn gezaag te willen luisteren en om samen te ontdekken hoe doctoreren werkt. Jannes, met je nuchtere blik slaag je erin om alles mooi in perspectief te plaatsen en je hulp bij MAI zal ik niet vlug vergeten. Vishvas and Raúl, I hope I didn't annoy you guys too much in our office. I'm still not sure whether those neutrinos really exist ;) Maarten, Wim, Nils, Bart, Sam, Nathalie, Kajetan, jullie waren toffe collega's. Ik zal steeds met plezier terugdenken aan de tweede verdieping van N3 en zijn bewoners.

Ook alle andere collegas van het INW hebben de vele lunch- en koffiepauzes mee gekleurd, alsook het dagelijks partijtje pingpong. Bedankt Sander, Matthias, Sam, Céline, Ianthe, Ward, Linda en al de rest, het was steevast gezellig op het INW.

Naast het INW mocht ik ook proeven van het leven als econoom. Het was aangenaam verpozen in kantoor 019-020 en omstreken. 'Stinos Ronsos', 'Samuelito', Stijn 'Pantalonos', Angelos 'el greco', Ruben 'el blanco', 'Brambo', bedankt voor de juiste mix van serieuze gesprekken en onzin uitkramen. Martine en Eva, bedankt om mij steeds de weg te wijzen doorheen de administratieve jungle van de ugent. Glenn, bedankt om Samuel, Stijn en mij bij te staan tijdens ons historisch project. Ilse, Jozefien, Désirée en Sümeyra, goed dat jullie er waren om wat tegengewicht te bieden aan alle mannenpraat. Veel succes ook aan de nieuwe lichting, Tom en Sam, en aan de oude garde: Matthijs, Ruben, Mustafa, Lennert, Karolien en Vicky.

Wie ik ook nog wil bedanken zijn: de mannen van de 'Fysicasa' om mijn jaren in Gent mee kleur te geven, de mannen van Poperinge voor het nodige vertier in het weekend, 'the animal spirits' voor wat muziek op het werk, Greg voor wat muziek naast het werk, Tim en 'Jacques' om te tonen hoe het eraan toegaat aan de KULeuven, Tarik for taking our upcoming project to the next level, Diego for keeping me company in Zurich, Atsushi for hosting me in Tokyo, the members of my PhD jury for being kind, Peter, Véronique, Milan en Simon om mijn thesisstudent te willen zijn, en de mensen bij de bank.

Als laatste bedank ik graag mijn twee zussen voor het geven van het goeie voorbeeld, zowel academisch als ernaast. De Bonduelles, Talabi's en Bassambi's zullen lang na deze vijf jaar een bron van plezier blijven. Ik bedank mijn ouders voor hun goeie zorgen en voor de vele kansen die ik kreeg. Maar mijn grootste geluk de afgelopen 28 jaar, is dat ik Ellen heb mogen ontmoeten. Ellen, je bent op alle vlakken een topwijf! :)

Gent, mei 2017
Benjamin Vandermarliere

Table of Contents

1 Introduction

The work presented in this dissertation can be roughly divided into an empirical part and a methodological part. Each chapter reports on the results of a separate research project. The different chapters address a variety of economic and methodological research questions. A constant theme running through the chapters is the methodological approach grounded in network theory and network methodology.

The empirical part of this dissertation reports the results of three detailed analyses of data for specific economic systems: the Russian interbank lending market, international trade during the 20th century, and exchange-traded funds. The constant themes in these chapters are the use of **network theory** and of **statistical distributions** with the aim of uncovering stylized facts in the relevant quantities of the systems under study. The methodological part of this dissertation comprises two studies of novel methods that aim at capturing and properly quantifying the time-evolution of networks. Chapters 5 and 6 contribute to the science of how propagation occurs on **temporal networks**. It does so by developing a novel temporal network representation and accompanying measures. Chapter 7 extends the existing literature on developing methodologies for the detection of changes in the large-scale structure of networks as they evolve over time.

In the forthcoming section of this introductory chapter we set the stage for the key concepts used throughout the thesis. In Section 1.2 we then sketch the outline of the upcoming chapters.

1.1 Introduction

1.1.1 Networks

Networks can be used to succinctly describe and improve our understanding of a whole range of complex systems packed with intricately interacting people or things [1]. Representing these people or things as nodes and their mutual interactions as edges in a network, enables one to disentangle and find order in what at first might look like a hotchpotch of connections. Over the past twenty years, this abstractive property of networks has proven its merit time and time again. The concept of a complex network has made its way to a wide variety of scientific disciplines, and lies at the heart of some of the most recent society-transforming technologies [2].

Biologists are putting network science to work with the aim of mapping the interactions of genes, proteins, metabolites and other cellular components [3]. These efforts have for example resulted in an improved understanding of genetic diseases [4]. In medicine, and more particular epidemiology, the spread of diseases is now better comprehended than ever before: network theory can nowadays, e.g., help pinpointing who to inoculate in order to prevent worse [5, 6]. Neuroscientists are charting the neuronal connections of mammalian brains [7], while engineers are using networks to understand the cascading failures of power grids [8, 9] or the congestion of roads [10].

Next to the above-mentioned applications in the physical sciences, network theory has also proved to be an enormous asset in the social sciences, where it actually originated. Indeed, complex network theory has yielded explanations for social phenomena in a variety of disciplines ranging from psychology to economics [11]. Its success is due to the fact that network tools can capture one of the most potent ideas in the social sciences: the notion that individuals are embedded in high-density webs of social relations and interactions. In 1934 networks first appeared under the form of "sociometry", a technique for eliciting and graphically representing individuals' subjective feelings toward each other [12]. From there on network theory advanced along several fronts up till today [11]. A first well-known discovery was the 'strength of weak ties' by Granovetter [13]. People with whom one has strong ties, tend to know each other well too. Hence, some of the information these people pass along is redundant. In contrast, people with whom one has a weak tie (e.g. an acquaintance), tend to be not connected to one's strong network and therefore are more likely to be sources of novel information that might get you a new job or boyfriend [11]. A second highly popularized concept is the 'small world' problem: upon randomly selecting two persons from a population, what are the chances that they would know each other, or more generally, how many people would be needed to link them? From a battery of experiments, it is now established that one can connect any two people in the world via at most six people,

a phenomenon that is commonly referred to as 'six degrees of separation' [14, 15].

One of the last social sciences to embrace network theory was economics [16–18]. Just as with the other fields where methodologies gleaned from network theory gained a foothold, economists initially used them to get a grasp on interaction data. Whether these are countries trading [19, 20], people paying [21, 22] or banks lending [23–25], all could already benefit from the complex systems and network perspective. Besides this descriptive work, the effort to include networks in theoretical models has also been revving up. They are appearing in models on economic development [26], production [27], systemic risk [28, 29], international trade relationships [30], and game theory in general [31, 32].

Network theory not only has had a huge impact on science. It has also affected how billions of people interact daily and has profoundly changed the business world too. The essence of game-changing companies like Facebook, LinkedIn and Twitter is building an all-encompassing social network and exploiting it. And ever since 1998, Google has been continuously mapping the WWW, a network of webpages and hyperlinks, in order to feed their search engine ranking algorithm grounded in network theory.

1.1.2 Network analysis in economics

One of the major goals of economics is to capture and understand how people, firms and countries interact e.g. through trade, payments or the extension of credit. A natural candidate to uncover patterns in this myriad of interactions is network theory. Since the advent of modern computers and the accompanying digitalization of the economy, all sorts of data on economic interactions has been consistently recorded. Representing these novel datasets as networks, provided researchers with a heretofore inaccessible way of addressing economic questions.

In general, a network analysis always follows the same pattern, irrespective of the system under study. After having decided on how to translate the raw data to a network representation, one tries to establish the relative importance of each node and edge. Is a given node a central player in the network? Is the edge crucial in keeping the network from disintegrating or is it redundant? The answers to those questions are often linked to the next and more challenging step in a network analysis, namely the endeavor to uncover non-trivial structures that lay hidden in the spaghetti of connections of the system. A network might, e.g., consist of two densely connected groups of nodes, with little connections between those two groups.

As an illustration, let us suppose we have data on the bilateral trade flows among a group of countries. It is obvious that the countries should be represented as nodes and that the edges somehow capture the details of the trade flows. Yet, the edges representing the trade flows can be constructed to mean different things. An

edge between two countries can signify there was some form of trade, irrespective of the amount or direction [33]. Edges can be directed to distinguish a situation where one country sells goods to another country but not vice versa [30, 34–37]. The edges can be weighted with the volume of exports [38–40], allowing us to distinguish countries that trade a lot. These weights can be further refined by making the edges proportional to either the total trade of the country [41] or to the GDP of the country [42]. Indeed, a trade flow of a million USD is of greater importance to a small economy that trades little, than to a large country that trades a lot. Most importantly, the edges can be enriched with a time stamp, such that the network representation of trade flows becomes dynamic. With those addition one can for example study the temporal trends and the system's response to shocks like war or depression. Deciding on the appropriate level of abstraction often depends as much on the envisioned research questions as on the quality of the data.

Once the network is constructed, we are able to assess the roles of certain nodes or edges of interest. Questions can be addressed like: Which countries are the major players in the world's markets? Do the major players change over times? Is the world economy generally becoming more or less connected over time? Is this the first era of globalisation or has an equivalent event occurred before? Does a country have an open or a closed economy? etc... Each of these questions can be addressed by a tool from the network scientist's toolbox. The appropriate tool, however, usually has to be adapted or reinterpreted for its specific use. Then, as a final step or in parallel, one can focus on any larger scale structures in the network. For this example on international trade, we can test to what extent we can detect emergent structures like trade blocs [43, 44], the latter being defined as a group of countries that on average trades more with fellow group members than with non-group members. Another example of a proposed structure in international trade results from the core-periphery hypothesis formulated by Wallerstein [45, 46]. He claims there is a core of developed countries that densely trade with each other, and next to these, there is a periphery of countries that trade with the core, but hardly trade among themselves. For more on the above, we refer the reader to Chapter 3 where we dive into the web of international trade and address exactly these questions in much more detail.

The main reason to illustrate the importance of network analysis in economics with an example from international trade data is that it was one of the very first applications of networks in economics. In the meantime, a whole range of economic systems have been investigated from the network perspective [16]. A famous example was the mapping of the architecture of the international corporate control ownership [47]. The authors of this study found that a large portion of control flows to a small and tightly-knit core of financial institutions. But other examples of economic systems studied via a network analyses are abound: the R&D collaboration [48] and board member overlap [49] of companies, migration between

countries [50], foreign direct investment flows [51], global supply chains [52], the bitcoin market [21], among others. In all of the above examples, networks add value by enabling us to see the forest for the trees.

This overview would not be complete without mentioning the recent research surge of studies on interbank credit networks. Ever since the highly interconnected interbank credit market acted as a catalyst for the complete meltdown of financial markets in the finacial crisis of 2008, the topic has received intense attention from academia [53]. Many empirical studies have investigated the topology of interbank markets like, e.g., those of Austria [54], The Netherlands [55] and Italy [56]. Several stylized features were discovered in the above cases: fat-tailed degree distributions where the highest connected banks in the tail are often too-big-to-fail [57], and a core-periphery structure where the core consists of intermediating banks [58]. These findings are then used as input for simulations of financial contagion. The goal of these simulations, in turn, is to understand the interplay of the interbank network topology and the dynamics of financial contagion [59]. Through the work presented in Chapter 2, we contribute to this field with an analysis of the interbank market of Russia during a period that covers two banking crises of a very different nature.

1.1.3 Temporal networks

The strength of adopting a network perspective lies in the ability to formally capture the relevant quantitative properties of a complex system so that one can better understand its collective behavior [60]. As mentioned in the previous section, the granularity of this abstraction will depend on the research question formulated. Often, e.g., adding information about the timing of interactions can be crucial to fully grasp the processes in the system under investigation [61]. Suppose one is studying the spread of an infectious disease via interpersonal contacts in a group of people. One approach is to simply map the implicit social network of the group and then use this as a basis for contagion simulations. More subtly, and arguably better, is to take into account just when or in which 'order' these people usually meet. Shaking hands with a person one month before he gets infected is a whole different thing than shaking hands the day his contagiousness is at its peak. Neglecting the aspect of timing by amalgamating the interpersonal contacts over a certain period into a 'flat' or static network, and performing contagion simulations on this static network, will yield different results than respecting the time ordering of the actual interpersonal contacts. The drawback with this often more accurate temporal approach, is that there are not so many methods available. Indeed, the development of appropriate methods to study temporal networks is a relatively young field, partly because those methods are more difficult to develop compared to what has been commonly done for characterizing static networks [60].

Many systems have been modeled as temporal networks: the proximity of humans [62] and animals [63]; human communication [64]; and travel and transportation [65]. Temporal networks are also used in the study of brain data [66] and of course in computer science [67], where the field originally emerged. In general, any system with pairwise interactions and information about time could be modeled as a temporal network. But, for such modeling to be practically useful, there should be some structure in both time and network topology that affects dynamic systems on the network. The latter include the spreading of infectious diseases, opinions, rumors, in social networks; information packets in computer networks; various types of signaling in biology, and more [60].

Contributers to the young field of temporal networks are still sorting out best practices. Thereby, one of the key issues is to identify the most appropriate representation of temporal interaction data given the research questions that are being addressed. Many often subtly different ways have been developed to represent and reason about temporal networks. For an excellent introduction to the various representations available in the literature, we refer the reader to the overview papers by Holme [60, 61]. A more straightforward way to deal with temporal networks is to present them as sequences of static networks [68]. Suppose we have interaction data time-stamped with a daily resolution spanning a year in total. Building 12 monthly networks with all the interactions of one month combined into one network, is an example of representing temporal interaction data as sequences of static networks. This is the most straightforward path to make the transition from static to temporal networks, as one can directly transfer the static network measures to this version of temporal networks. Calculate the measure for each time slice, and the evolution over time of that measure tells us something about the evolution of the temporal network. But, for this representation to make sense, the temporal resolution should be low. If the resolution is too high, and thus one coarse-grains too much, then network sequences are often not a good idea.

The two empirical chapters in this thesis that study the evolution of a network over time, both use the above-mentioned temporal network representation. For the interbank network in Chapter 2 this is vindicated as the interbank activity has a strong monthly periodicity. For the trade network studied in Chapter 3 the trade flows can be argued to evolve slower than the yearly timescale. Also the methodological Chapter 7 works with this representation of temporal networks, as it builds on an earlier developed framework that uses sequences of static networks.

For the other two methodological Chapters 5 and 6, we demonstrate the use of another version of temporal networks and related measures called "temporal webs"[1] to capture and analyze a variety of problems. Temporal webs are distinguished by their use of cross-temporal interaction and/or inheritance links. So,

[1] Other researchers, working in parallel on this version of temporal networks coined this representation "time-node networks" [60, 69–71].

rather than being a sequence of network time slices connected by node membership, they are always monolithic graphs of the interaction structures across time. A pure temporal web has only cross-temporal edges to create a single acyclic directed graph. This construction has certain advantages in communication networks for which transmission and reception may take several time steps. It also embodies some specific advantages for analysis through the availability of approaches that work on large, sparse directed adjacency matrices.

1.1.4 Statistical analysis of distributions

Next to the common theme of networks running through all but one chapter, Chapters 2 and 4 are each built around a detailed statistical analysis of probability distributions for relevant quantities in complex economic systems. In Chapter 2, a statistical analysis is used to search for the theoretical distributions that best describe a range of measures related to the Russian interbank network. In Chapter 4, the size distribution of exchange-traded funds is subject to an investigation into a presumable log-periodic structure decorating the distribution.

The remainder of this section is structured as follows. Subsection 1.1.4.1 discusses how the statistical analysis of distributions played an important role in network science, and relates this to the work on the Russian interbank network of Chapter 2. Subsection 1.1.4.2 introduces the concept of discrete scale invariance and discusses its link with Chapter 4 concerning the exchange-traded funds.

1.1.4.1 Scale-free networks

During the early years of network science, the analysis of statistical distributions played a key role as it was central to the detection of scale-free networks. Scale-free networks are networks with a degree distribution that can be (partly) described by a power law. In many circumstances this is the situation in the tail part of the distribution. The only way to make sure that one was dealing with a power-law degree distribution, was through a detailed statistical analysis of the degree distribution [72].

A degree distribution that satisfies a power law implies, simply stated, that most nodes will have barely any connections, some nodes will have few connections, and very few nodes will be connected to almost all other nodes. These highly-connected nodes can have hundreds, thousandths or even millions of links to others [73]. This stands in stark contrast to the many empirical quantities that cluster around a typical average value like the height of people or the speed of cars. All of these have somewhat varying values, but their distributions place a negligible amount of probability far from the typical value, making the typical value representative of most observations [74]. In this sense, scale-free networks appear to have no characteristic scale.

This concept of a scale-free network might at first seem something rather exotic. But as network scientists were exploring ever new data sets in the early years, scale-free networks started to show up in a wide range of complex systems, independent of their age, function, and scope. Internet routers were found to be connected as a scale-free network [75], just as the web pages of the WWW are by URLs [76]. The way actors co-appear in movies, scientists co-author, or people have sexual contacts, are other examples [74]. Also in biology, metabolic molecule interactions or cell-regulating proteins are linked as a scale-free network [72].

Hence, as soon as interbank networks started to receive some attention in the wake of the 2008 financial crisis, researchers also tested these interbank networks for scale-freeness. The first to supposedly report a scale-free interbank network were Boss et al. for Austria [54] and over time similar claims were also made for Brazil [77], Italy [78], and the European market [79]. Other studies painted a more nuanced picture [53] and provided evidence that the degree distribution of interbank networks is by no means a power law at low values of the degrees, but are characterized by so-called fat tails [56, 80]. The initial goal of our work on the Russian interbank network, presented in Chapter 2, was to contribute to this discussion by performing a thorough statistical analysis of the degree distribution via the method described in Ref. [74]. Especially because it was evident from interbank default contagion simulations that the degree distribution was a crucial parameter in grasping the ramifications of a default event [53, 59]. As we progressed with the analysis, we expanded it to include additional interbank network quantities, which we identified of importance upon designing detailed simulations of the impact of default events. In general, we found no strong evidence to corroborate the power-law hypothesis. It became evident, however, that the distributions of interbank measures are systematically fat tailed.

1.1.4.2 Discrete scale invariance

Not only the wiring schemes of networks can be scale-free (also called scale-invariant). Other, non-network related, quantities have been found to have a power-law signature in their distributions. Examples include the intensities of earthquakes (Gutenberg-Richter law) [81] and wars [82], the frequency of occurrence of unique words in a novel [82], or the sales volume of bestselling books [83]. Also in economics, power-law distributions conquered their spot in the limelight, as one found that the asymptotic part of income [84], of the net worth of rich individuals [82], and of the size of companies [85] are scale-free distributed. In finance, power laws were found to govern the distributions of relevant financial fluctuations, such as fluctuations in stock price [86], trading volume [87] and the number of trades [88]. This means that large fluctuations, up- and certainly downward ones, are more likely than one would expect using e.g. standard (log)normal models of financial returns. Recognizing this fact is quintessential when trading-off risk and return in

financial markets. The collapse of the hedge fund management firm Long-Term Capital Management (LTCM) is just a case in point [89].

Because of this pervasiveness of power laws in finance and economics, one of the first things we did was to test for scale invariance while exploring the data on exchange-traded funds (ETFs) in the preliminary work leading up to Chapter 4. The tests indicated that we were dealing with log-normal instead of power-law behavior. However we did notice that the size distribution was decorated with a log-periodic structure, a tell-tale sign of *discrete* scale invariance (DSI). A system with the DSI property, has no distinct scale just as is the case with a scale-invariant system. The difference between the continuous and discrete version lies in the existence of a preferred scaling ratio for the latter [90]. In other words, rather than a single or continuous spectrum of ETF sizes, ETFs form groups of preferred sizes clustered around values organized in a geometrical series. DSI is connected to the concept of fractals and has been found in a range of complex systems often linked to a notion of hierarchy. One example is the way groups of people are organized in society [91, 92] which can in turn be related to the DSI found in the herding behavior in stock markets [93]. Other examples of DSI are the size of eddies in turbulence [94], the length of ruptures during earthquakes [95], and the diameter of bronchial branches [96],

In Chapter 4 we use a detailed statistical analysis to test whether the size distributions of ETFs really display the DSI property. Once this is established, we offer some possible generative mechanisms and take a closer look at the composition and properties of the emergent discrete scales.

1.2 Outline

This PhD dissertation consists of an empirical and a methodological part. The empirical part covers three analyses of economic systems: the Russian interbank lending market, international trade during the long 20th century, and exchange-traded funds. The methodological part collates three contributions to temporal network theory. Chapters 5 and 6 develop the concept of "temporal webs". The Chapter 7 deals with the further development of efficient techniques that aim at detecting changes in the large-scale structure of networks as they evolve over time.

Part I: Stylized collective properties of interbank networks, historical trade and exchange-traded funds

Chapter 2: Analysis of the Russian interbank lending system For this first empirical chapter we performed a network analysis of the Russian interbank lending system. The chapter is co-authored by Alexei Karas, Jan Ryckebusch and Koen Schoors and has been published as a paper in "Physica A: Statistical mechanics and

its applications" under the title "Beyond the power law: Uncovering stylized facts in interbank networks" [25]. The motivation of this work was to gather stylized facts in the properties of interbank networks. Those properties can subsequently be used as input for interbank network simulations. The main focus lies on finding the right distribution to describe a range of interbank network measures, and this in times of normal operation periods as well as in crisis periods.

We analyzed daily data on bilateral interbank exposures and monthly bank balance sheets of the Russian interbank market over August 1998 - October 2004. Specifically, we examined the distributions of (un)directed (un)weighted degree, nodal attributes (bank assets, capital and capital-to-assets ratio) and edge weights (loan size and counterparty exposure). We searched for the theoretical distribution that fits the data best and report the "best" fit parameters. We observed that all studied distributions are heavy tailed, where the fat tail typically contains 20% of the data and can be systematically described by a truncated power law. In most cases, however, separating the bulk and tail parts of the data is hard, so we proceeded to study the full range of the events. We found that the stretched exponential and the log-normal distributions fit the full range of the data best and that these conclusions are robust to (1) whether we aggregate the data over a week, month, quarter or year; (2) whether we look at the "growth" versus "maturity" phases of interbank market development; and (3) with minor exceptions, whether we look at the "normal" versus "crisis" operation periods. In line with prior research, we found that the network topology changes substantially as the interbank market moves from a "normal" to a "crisis" operation period.

Chapter 3: Structure and evolution of the world's historical trade patterns
The work presented here is a product of a broader research project co-developed with Samuel Standaert and Stijn Ronsse[2]. For this research line, we combined historical data with network theory and advanced statistical methods, in order to address economic historical questions on international trade. This chapter in particular, studies trade patterns which emerged during the long 20th century. It will be published in 2017 as a chapter in the book "Networks of International Trade and Investment" edited by Sara Gorgoni, Alessia Amighini and Matthew Smith.

This chapter studies the structure of the worldwide trade network from the 1880s to the late 1980s, looking specifically for patterns corresponding to globalization, regionalization or a core-periphery structure. After constructing the historical trade integration network, we use temporal stochastic block models (SBMs) to extract the meso-scale network structure. This SBM methodology makes full use of all available data, takes the time dimension into account and does not make

[2]The other products of this research project are a paper titled "Historical trade integration: globalization and the distance puzzle in the long twentieth century" published in "Cliometrica" [20] and a popularized synopsis chapter called "Groeipolen en macht: het internationale handelsnetwerk in de lange 20ste eeuw" published in the book "De hermaakbare wereld? Essays over globalisering" [97].

a priori assumptions about the structure of the network. The total time period was divided into three time intervals: the first globalization wave (1880-1913), the Interbellum, a period of de-globalization (1919-1939) and the second globalization wave until the fall of the Berlin wall (1946-1989). During the first globalization wave, we find a very strong core-periphery structure, where the core is made up of countries like the USA and Great Britain. However, over the next two periods, we see the slow dismantlement of this structure into a hub-and-spoke pattern where regional clusters are linked to a central hub, a sign of increasing regionalization.

Chapter 4: Discrete hierarchy of sizes and performances in the exchange-traded fund universe The fourth and final empirical chapter studies exchange-traded funds (ETFs) and is joint work with Peter Cauwels, Koen Schoors, Jan Ryckebusch and Didier Sornette. This chapter is published as a paper called "Discrete hierarchy of sizes and performances in the exchange-traded fund universe" in "Physica A: Statistical mechanics and its applications" [98].

The chapter reports a detailed statistical analyses of the size distribution of a universe of ETFs. This analysis uncovered a discrete hierarchy of sizes, which imprints a log-periodic structure on the probability distribution of ETF sizes that dominates the details of the asymptotic tail. This allows us to propose a classification of the studied universe of ETFs into seven size layers approximately organized according to a multiplicative ratio of 3.5 in their total market capitalization. Introducing a similarity metric generalising the Herfindahl index, we find that the largest ETFs exhibit a significantly stronger intra-layer and interlayer similarity compared with the smaller ETFs. Comparing the performance across the seven discerned ETF size layers, we find an inverse size effect, namely large ETFs perform significantly better than the small ones both in 2014 and 2015.

Part II: Methodological contributions to temporal network theory

Chapters 5 and 6: Temporal webs The next two chapters cover the work on temporal networks which was mainly a joint effort with Aaron Bramson. In these two chapters we propose a framework to study transmission on temporal networks. Chapter five introduces the concept and develops a first set of new measures. Chapter six elaborates on the technique and improves the most promising measure covered in the preceding chapter. Both chapters have been published as papers. The first one "Dynamical properties of interaction data" has been published in the "Journal of Complex Networks" [99], the follow-up paper "Benchmarking measures of network influence" in "Scientific Reports" [100].

Identifying key agents for the transmission of diseases (ideas, technology, etc.) across networks has predominantly relied on measures of centrality on a static

base network or a temporally flattened graph of agent interactions. The proposed approach examines the dynamical properties of transmission via novel measures on an integrated, temporally extended network representation of interaction data across time. Because it encodes time and interactions as network connections, static network measures can be applied to this 'temporal web' to reveal features of the dynamics themselves. In chapter five we introduce the technique, provide the technical details and apply it to prototypical agent-based implementations of the well-known SEIR and SEIS epidemiological models. In the follow-up chapter six we further develop the previously introduced temporal knockout (TKO) score and argue that this TKO score is an effective benchmark measure for evaluating the accuracy of other, often more practical, measures of influence. Benchmarked to the traditional network measures applied to the induced flat graphs, we find that none of them are accurate predictors of network propagation influence, expressed by TKO, on the systems studied. However, temporal networks and the TKO measure provide the requisite targets for the search for effective predictive measures.

Chapter 7: Change-point detection in temporal networks The final chapter of this dissertation is joint work with Simon de Ridder and Jan Ryckebusch. Here, we worked with the more traditional representation of a temporal network as a sequence of static networks. From this vantage point, we extend a framework to detect significant structural changes in networks as they evolve over time. This chapter has been published as a paper titled "Detection and localization of change points in temporal networks with the aid of stochastic block models" in the "Journal of Statistical Mechanics: Theory and Experiment" [101].

We extended the proposed change-point detection methodology of [102] to also include the versatile stochastic block models (SBMs) as a parametric family for reconstructing the empirical networks. In this chapter five different techniques for change point detection are compared on prototypical temporal networks, including empirical and synthetic ones. We find that none of the considered methods can consistently outperform the others when it comes to detecting and locating the expected change points in empirical temporal networks. With respect to the precision and the recall of the results of the change points, we find that the method based on a degree-corrected SBM has better recall properties than other dedicated methods, especially for sparse networks and smaller sliding time window widths.

Computational setup

The scripts to analyze the data in the first part of this dissertation were coded in python. The analyses were all performed on a regular laptop (Intel(R) Core(TM) i7-5600U CPU 2.60GHz $\times 4$ running ubuntu). Chapter 2 made extensive use of the power law package [103] and the entire pipeline took about 15 minutes. Chapter 3

in its turn used the graph-tool library [104] where fitting a block model for one time period on average took 5 minutes. This is the same order of magnitude as the analysis of the ETFs in Chapter 4.

The code for the numerical simulations in Chapters 5 and 6 was written in python too and is built with elements from the networkx library [105]. These were run on a cluster of 16 cores, where running the entire pipeline took about 24 hours. We acknowledge that python is not the optimal language for this kind of numerical simulations, hence the next iteration of the temporal web framework is currently being coded in c++ by Kevin Hoefman. For the final chapter the code was written in c++ by Simon de Ridder and ran on a regular office desktop with 4 processors. One analysis of a temporal network, including the bootstrapping to ascertain the significance of a proposed change point, takes about 14 days.

References

[1] Réka Albert and Albert-László Barabási. *Statistical mechanics of complex networks*. Reviews of modern physics, 74(1):47, 2002.

[2] Albert-László Barabási. *Network science*. Cambridge University Press, 2016.

[3] Uri Alon. *Biological networks: the tinkerer as an engineer*. Science, 301(5641):1866–1867, 2003.

[4] Albert-László Barabási, Natali Gulbahce, and Joseph Loscalzo. *Network medicine: a network-based approach to human disease*. Nature Reviews Genetics, 12(1):56–68, 2011.

[5] Leon Danon, Ashley P Ford, Thomas House, Chris P Jewell, Matt J Keeling, Gareth O Roberts, Joshua V Ross, and Matthew C Vernon. *Networks and the epidemiology of infectious disease*. Interdisciplinary perspectives on infectious diseases, 2011, 2011.

[6] Dirk Brockmann and Dirk Helbing. *The hidden geometry of complex, network-driven contagion phenomena*. Science, 342(6164):1337–1342, 2013.

[7] Olaf Sporns, Giulio Tononi, and Rolf Kötter. *The human connectome: a structural description of the human brain*. PLoS Comput Biol, 1(4):e42, 2005.

[8] Réka Albert, István Albert, and Gary L Nakarado. *Structural vulnerability of the North American power grid*. Physical review E, 69(2):025103, 2004.

[9] Giuliano Andrea Pagani and Marco Aiello. *The power grid as a complex network: a survey*. Physica A: Statistical Mechanics and its Applications, 392(11):2688–2700, 2013.

[10] Ming Hua and Jian Pei. *Probabilistic path queries in road networks: traffic uncertainty aware path selection*. In Proceedings of the 13th International Conference on Extending Database Technology, pages 347–358. ACM, 2010.

[11] Stephen P Borgatti, Ajay Mehra, Daniel J Brass, and Giuseppe Labianca. *Network analysis in the social sciences*. science, 323(5916):892–895, 2009.

[12] Jacob Levy Moreno. *Who shall survive?: A new approach to the problem of human interrelations.* 1934.

[13] Mark S Granovetter. *The strength of weak ties*. American journal of sociology, 78(6):1360–1380, 1973.

[14] Jeffrey Travers and Stanley Milgram. *The small world problem*. Phychology Today, 1:61–67, 1967.

[15] Jeffrey Travers and Stanley Milgram. *An experimental study of the small world problem*. Sociometry, pages 425–443, 1969.

[16] Frank Schweitzer, Giorgio Fagiolo, Didier Sornette, Fernando Vega-Redondo, Alessandro Vespignani, and Douglas R White. *Economic networks: The new challenges*. science, 325(5939):422–425, 2009.

[17] Matthew O Jackson. *Social and economic networks*. Princeton university press, 2010.

[18] David Easley and Jon Kleinberg. *Networks, crowds, and markets: Reasoning about a highly connected world*. Cambridge University Press, 2010.

[19] Thomas Chaney. *The network structure of international trade*. The American Economic Review, 104(11):3600–3634, 2014.

[20] Samuel Standaert, Stijn Ronsse, and Benjamin Vandermarliere. *Historical trade integration: globalization and the distance puzzle in the long twentieth century*. Cliometrica, 10(2):225–250, 2016.

[21] Dániel Kondor, Márton Pósfai, István Csabai, and Gábor Vattay. *Do the rich get richer? An empirical analysis of the Bitcoin transaction network*. PloS one, 9(2):e86197, 2014.

[22] Yves-Alexandre De Montjoye, Laura Radaelli, Vivek Kumar Singh, et al. *Unique in the shopping mall: On the reidentifiability of credit card metadata*. Science, 347(6221):536–539, 2015.

[23] Sam Langfield, Zijun Liu, and Tomohiro Ota. *Mapping the UK interbank system*. Journal of Banking & Finance, 45:288–303, 2014.

[24] Camelia Minoiu and Javier A Reyes. *A network analysis of global banking: 1978–2010*. Journal of Financial Stability, 9(2):168–184, 2013.

[25] Benjamin Vandermarliere, Alexei Karas, Jan Ryckebusch, and Koen Schoors. *Beyond the power law: Uncovering stylized facts in interbank networks*. Physica A: Statistical Mechanics and its Applications, 428:443–457, 2015.

[26] Daron Acemoglu, Camilo Garcia-Jimeno, and James A Robinson. *State capacity and economic development: A network approach*. The American Economic Review, 105(8):2364–2409, 2015.

[27] Vasco M Carvalho. *From micro to macro via production networks*. The Journal of Economic Perspectives, 28(4):23–47, 2014.

[28] Prasanna Gai and Sujit Kapadia. *Contagion in financial networks*. In Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences, page rspa20090410. The Royal Society, 2010.

[29] Stefano Battiston, Guido Caldarelli, Robert May, Tarik Roukny, and Joseph E Stiglitz. *The price of complexity in financial networks*. Proceedings of the National Academy of Sciences, 113(36):10031–10036, 2016.

[30] Raja Kali and Javier Reyes. *The architecture of globalization: a network approach to international economic integration*. Journal of International Business Studies, 38(4):595–620, 2007.

[31] Andrea Galeotti, Sanjeev Goyal, Matthew O Jackson, Fernando Vega-Redondo, and Leeat Yariv. *Network games*. The review of economic studies, 77(1):218–244, 2010.

[32] Sanjeev Goyal, Hoda Heidari, and Michael Kearns. *Competitive contagion in networks*. Games and Economic Behavior, 2014.

[33] Tiziano Squartini, Giorgio Fagiolo, and Diego Garlaschelli. *Randomizing world trade. I. A binary network analysis*. Phys. Rev. E, 84:046117, Oct 2011.

[34] Ma Ángeles Serrano and Marián Boguñá. *Topology of the world trade web*. Physical Review E, 68(1):015101, 2003.

[35] Xiang Li, Yu Ying Jin, and Guanrong Chen. *Complexity and synchronization of the world trade web*. Physica A: Statistical Mechanics and its Applications, 328(1):287–296, 2003.

[36] Diego Garlaschelli and Maria I Loffredo. *Fitness-dependent topological properties of the world trade web*. Physical review letters, 93(18):188701, 2004.

[37] Diego Garlaschelli and Maria I Loffredo. *Structure and evolution of the world trade network*. Physica A: Statistical Mechanics and its Applications, 355(1):138–144, 2005.

[38] Giorgio Fagiolo, Javier Reyes, and Stefano Schiavo. *On the topological properties of the world trade web: A weighted network analysis*. Physica A: Statistical Mechanics and its Applications, 387(15):3868–3873, 2008.

[39] Giorgio Fagiolo, Javier Reyes, and Stefano Schiavo. *World-trade web: Topological properties, dynamics, and evolution*. Physical Review E, 79(3):036115, 2009.

[40] Luca De Benedictis and Lucia Tajoli. *The world trade network*. The World Economy, 34(8):1417–1454, 2011.

[41] Irena Tzekina, Karan Danthi, and Daniel N Rockmore. *Evolution of community structure in the world trade web*. The European Physical Journal B, 63(4):541–545, 2008.

[42] Gautier M Krings, Jean-François Carpantier, and Jean-Charles Delvenne. *Trade integration and trade imbalances in the European Union: a network perspective*. PloS one, 9(1):e83448, 2014.

[43] Matteo Barigozzi, Giorgio Fagiolo, and Giuseppe Mangioni. *Identifying the community structure of the international-trade multi-network*. Physica A: statistical mechanics and its applications, 390(11):2051–2066, 2011.

[44] Javier Reyes, Rossitza Wooster, and Stuart Shirrell. *Regional trade agreements and the pattern of trade: A networks approach*. The World Economy, 37(8):1128–1151, 2014.

[45] Immanuel Wallerstein. *The rise and future demise of the world capitalist system: concepts for comparative analysis*. Comparative studies in society and history, 16(04):387–415, 1974.

[46] Immanuel Maurice Wallerstein. *World-systems analysis: An introduction*. Duke University Press, 2004.

[47] Stefania Vitali, James B Glattfelder, and Stefano Battiston. *The network of global corporate control*. PloS one, 6(10):e25995, 2011.

[48] Michael D König, Stefano Battiston, Mauro Napoletano, and Frank Schweitzer. *The efficiency and stability of R&D networks*. Games and Economic Behavior, 75(2):694–713, 2012.

[49] Dawn A Harris and Constance E Helfat. *The Board of Directors as a Social Network A New Perspective*. Journal of Management Inquiry, 16(3):228–237, 2007.

[50] Kyle F Davis, Paolo D'Odorico, Francesco Laio, and Luca Ridolfi. *Global spatio-temporal patterns in human migration: a complex network perspective*. PLoS One, 8(1):e53723, 2013.

[51] Homin Chen and Tain-Jy Chen. *Network linkages and location choice in foreign direct investment*. Journal of international business studies, 29(3):445–467, 1998.

[52] Morris A Cohen and Suman Mallik. *Global supply chains: research and applications*. Production and Operations Management, 6(3):193–210, 1997.

[53] Anne-Caroline Hüser. *Too interconnected to fail: A survey of the interbank networks literature*. 2015.

[54] Michael Boss, Helmut Elsinger, Martin Summer, and Stefan Thurner 4. *Network topology of the interbank market*. Quantitative Finance, 4(6):677–684, 2004.

[55] Iman van Lelyveld et al. *Finding the core: Network structure in interbank markets*. Journal of Banking & Finance, 49:27–40, 2014.

[56] Daniel Fricke and Thomas Lux. *On the distribution of links in the interbank network: Evidence from the e-mid overnight money market*. Empirical Economics, 49(4):1463–1495, 2015.

[57] Stefano Battiston, Michelangelo Puliga, Rahul Kaushik, Paolo Tasca, and Guido Caldarelli. *Debtrank: Too central to fail? financial networks, the fed and systemic risk*. Scientific reports, 2:541, 2012.

[58] Daniel Fricke and Thomas Lux. *Core–periphery structure in the overnight money market: evidence from the e-mid trading platform*. Computational Economics, 45(3):359–395, 2015.

[59] Tarik Roukny, Hugues Bersini, Hugues Pirotte, Guido Caldarelli, and Stefano Battiston. *Default cascades in complex networks: Topology and systemic risk*. Scientific reports, 3:2759, 2013.

[60] Petter Holme. *Modern temporal network theory: a colloquium*. The European Physical Journal B, 88(9):1–30, 2015.

[61] Petter Holme and Jari Saramäki. *Temporal Networks*. Physics Reports, 519(3):97–125, 2011.

[62] Yi-Qing Zhang, Xiang Li, Jian Xu, and Athanasios V Vasilakos. *Human interactive patterns in temporal networks*. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 45(2):214–222, 2015.

[63] M Carolyn Gates and Mark EJ Woolhouse. *Controlling infectious disease through the targeted manipulation of contact network structure*. Epidemics, 12:11–19, 2015.

[64] Jari Saramäki and Esteban Moro. *From seconds to months: an overview of multi-scale dynamics of mobile telephone calls*. The European Physical Journal B, 88(6):164, 2015.

[65] Martin Rosvall, Alcides V Esquivel, Andrea Lancichinetti, Jevin D West, and Renaud Lambiotte. *Memory in network flows and its effects on spreading dynamics and community detection*. Nature communications, 5, 2014.

[66] Danielle S Bassett, Nicholas F Wymbs, M Puck Rombach, Mason A Porter, Peter J Mucha, and Scott T Grafton. *Task-based core-periphery organization of human brain dynamics*. PLoS Comput Biol, 9(9):e1003171, 2013.

[67] Rina Dechter, Itay Meiri, and Judea Pearl. *Temporal constraint networks*. Artificial intelligence, 49(1-3):61–95, 1991.

[68] Othon Michail. *An introduction to temporal graphs: An algorithmic perspective*. Internet Mathematics, 12(4):239–280, 2016.

[69] René Pfitzner, Ingo Scholtes, Antonios Garas, Claudio J Tessone, and Frank Schweitzer. *Betweenness preference: Quantifying correlations in the topological dynamics of temporal networks*. Physical review letters, 110(19):198701, 2013.

[70] Leo Speidel, Taro Takaguchi, and Naoki Masuda. *Community detection in directed acyclic graphs*. arXiv preprint arXiv:1503.05641, 2015.

[71] Taro Takaguchi, Yosuke Yano, and Yuichi Yoshida. *Coverage centralities for temporal networks*. The European Physical Journal B, 89(2):1–11, 2016.

[72] Albert-László Barabási. *Scale-free networks: a decade and beyond*. science, 325(5939):412–413, 2009.

[73] AL Barabasi and E. Bonabeau. *Scale-Free Networks*. Scientific American, 2003.

[74] Aaron Clauset, Cosma Rohilla Shalizi, and Mark EJ Newman. *Power-law distributions in empirical data*. SIAM review, 51(4):661–703, 2009.

[75] Réka Albert, Hawoong Jeong, and Albert-László Barabási. *Error and attack tolerance of complex networks*. nature, 406(6794):378–382, 2000.

[76] Albert-László Barabási, Réka Albert, and Hawoong Jeong. *Scale-free characteristics of random networks: the topology of the world-wide web*. Physica A: statistical mechanics and its applications, 281(1):69–77, 2000.

[77] Rama Cont, Amal Moussa, et al. *Network structure and systemic risk in banking systems*. 2010.

[78] Giulia De Masi, Giulia Iori, and Guido Caldarelli. *Fitness model for the Italian interbank money market*. Physical Review E, 74(6):066112, 2006.

[79] Ivan Alves, S Ferrari, P Franchini, JC Heam, P Jurca, S Langfield, S Laviola, F Liedorp, A Sánchez, S Tavolaro, et al. *The structure and resilience of the European interbank market*. Occasional Papers, 3, 2013.

[80] Giulia Iori, Giulia De Masi, Ovidiu Vasile Precup, Giampaolo Gabbi, and Guido Caldarelli. *A network analysis of the Italian overnight money market*. Journal of Economic Dynamics and Control, 32(1):259–278, 2008.

[81] Per Bak and Chao Tang. *Earthquakes as a self-organized critical phenomenon*. J. geophys. Res, 94(15):635–15, 1989.

[82] Mark EJ Newman. *Power laws, Pareto distributions and Zipf's law*. Contemporary physics, 46(5):323–351, 2005.

[83] A. P. Hackett. *70 Years of Best Sellers, 1896-1965*. R. R. Bowker Company, New York, 1967.

[84] Vilfredo Pareto. *Cours d'économie politique*, volume 1. Librairie Droz, 1964.

[85] Robert L Axtell. *Zipf distribution of US firm sizes*. Science, 293(5536):1818–1820, 2001.

[86] Vasiliki Plerou, Parameswaran Gopikrishnan, Luis A Nunes Amaral, Martin Meyer, and H Eugene Stanley. *Scaling of the distribution of price fluctuations of individual companies*. Physical review e, 60(6):6519, 1999.

[87] Parameswaran Gopikrishnan, Vasiliki Plerou, Luis A Nunes Amaral, Martin Meyer, and H Eugene Stanley. *Scaling of the distribution of fluctuations of financial market indices*. Physical Review E, 60(5):5305, 1999.

[88] Parameswaran Gopikrishnan, Vasiliki Plerou, Xavier Gabaix, and H Eugene Stanley. *Statistical properties of share volume traded in financial markets*. Physical Review E, 62(4):R4493, 2000.

[89] Roger Lowenstein. *When genius failed: the rise and fall of Long-Term Capital Management*. Random House trade paperbacks, 2000.

[90] Didier Sornette. *Discrete-scale invariance and complex dimensions*. Physics reports, 297(5):239–270, 1998.

[91] W-X Zhou, Dider Sornette, Russell A Hill, and Robin IM Dunbar. *Discrete hierarchical organization of social group sizes*. Proceedings of the Royal Society of London B: Biological Sciences, 272(1561):439–444, 2005.

[92] Benedikt Fuchs, Didier Sornette, and Stefan Thurner. *Fractal multi-level organisation of human groups in a virtual world*. Scientific reports, 4, 2014.

[93] W-X Zhou and Didier Sornette. *Self-organizing Ising model of financial markets*. The European Physical Journal B, 55(2):175–181, 2007.

[94] EA Novikov. *The effects of intermittency on statistical characteristics of turbulence and scale similarity of breakdown coefficients*. Physics of Fluids A: Fluid Dynamics, 2(5):814–820, 1990.

[95] A Sornette and D Sornette. *Self-organized criticality and earthquakes*. EPL (Europhysics Letters), 9(3):197, 1989.

[96] Bruce J West. *Physiology in fractal dimensions: error tolerance*. Annals of biomedical engineering, 18(2):135–149, 1990.

[97] Julie Carlier, Christopher Parker, and Eric Vanhaute. *De hermaakbare wereld? Essays over globalisering*. Gingko, 2016.

[98] Benjamin Vandermarliere, Jan Ryckebusch, Koen Schoors, Peter Cauwels, and Didier Sornette. *Discrete hierarchy of sizes and performances in the exchange-traded fund universe*. Physica A: Statistical Mechanics and its Applications, 469:111–123, 2017.

[99] Aaron Bramson and Benjamin Vandermarliere. *Dynamical properties of interaction data*. Journal of Complex Networks, 4(1):87–114, 2016.

[100] Aaron Bramson and Benjamin Vandermarliere. *Benchmarking measures of network influence*. Scientific Reports, 6, 2016.

[101] Simon De Ridder, Benjamin Vandermarliere, and Jan Ryckebusch. *Detection and localization of change points in temporal networks with the aid of stochastic block models*. Journal of Statistical Mechanics: Theory and Experiment, 2016(11):113302, 2016.

[102] Leto Peel and Aaron Clauset. *Detecting Change Points in the Large-Scale Structure of Evolving Networks*. In Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence, 2015.

[103] Jeff Alstott, Ed Bullmore, and Dietmar Plenz. *powerlaw: a Python package for analysis of heavy-tailed distributions*. PloS one, 9(1):e85777, 2014.

[104] Tiago P Peixoto. *The graph-tool python library*. figshare, 2014.

[105] Daniel A Schult and P Swart. *Exploring network structure, dynamics, and function using NetworkX*. In Proceedings of the 7th Python in Science Conferences (SciPy 2008), volume 2008, pages 11–16, 2008.

Part I

Stylized collective properties of interbank networks, historical trade and exchange-traded funds

2

Beyond the power law: Uncovering stylized facts in interbank networks

2.1 Abstract

We use daily data on bilateral interbank exposures and monthly bank balance sheets to study network characteristics of the Russian interbank market over Aug 1998 - Oct 2004. Specifically, we examine the distributions of (un)directed (un) weighted degree, nodal attributes (bank assets, capital and capital-to-assets ratio) and edge weights (loan size and counterparty exposure). We search for the theoretical distribution that fits the data best and report the "best" fit parameters.

We observe that all studied distributions are heavy tailed. The fat tail typically contains 20% of the data and can be mostly described well by a truncated power law. Also the power law, stretched exponential and log-normal provide reasonably good fits to the tails of the data. In most cases, however, separating the bulk and tail parts of the data is hard, so we proceed to study the full range of the events. We find that the stretched exponential and the log-normal distributions fit the full range of the data best. These conclusions are robust to 1) whether we aggregate the data over a week, month, quarter or year; 2) whether we look at the "growth" versus "maturity" phases of interbank market development; and 3) with minor exceptions, whether we look at the "normal" versus "crisis" operation periods. In line with prior research, we find that the network topology changes greatly as the interbank market moves from a "normal" to a "crisis" operation period.

2.2 Introduction

The frequency of an event follows a power law when that frequency varies as a power of some attribute of that event (e.g. its size). Power-law distributions have been claimed to occur in an extraordinarily diverse range of phenomena from the sizes of wars, earthquakes and computer files to the numbers of papers scientists write and citations those papers receive [1]. In economics and finance, power laws have been documented for income and wealth [2], the size of cities and firms, stock market returns, trading volume, international trade, and executive pay [3].

Most relevant to this paper, the tail parts of interbank network characteristics, such as degree distribution, have been shown to follow a power law too (see [4] for Brazil, [5, 6] for Austria, [7] for Japan, and [8] for the commercial banks in the US). This ubiquituous presence of power laws has resulted in an extensive search for universal dynamics that can explain their existence (see [9, 10] for examples of such search in interbank networks).

Recently, however, Clauset et. al. [11] (followed by [12]) call these findings into question. In particular, they criticize the commonly used methods for analyzing power-law data, such as least-squares fitting, which can produce inaccurate estimates of parameters for power-law distributions or provide no indication of whether the data obey a power law at all. Clauset et. al. propose a statistical framework for discerning and quantifying power-law behavior in empirical data and apply that framework to twenty-four real-world data sets, each of which has been conjectured to follow a power law. For most datasets they find moderate to weak evidence in favor of power laws.

This debate about the potential of power laws to capture the underlying network dynamics is important for economic policy. For example, during the recent financial crisis, the interbank lending market was one of the most important channels of financial contagion. The malfunctioning of the interconnectivity of the interbank lending network, caused a liquidity drought with consequences reverberating throughout the entire economy. Since then, interbank markets research has proliferated. In those studies one wishes to uncover the network topology of interbank markets, to understand how they function, and how they could catalyse a systemic meltdown [13, 14]. Current research on contagion in interbank markets often relies on a scale-free topology to simulate the interbank network [15, 16]. This choice likely affects the outcome of conducted stress tests (as is explicitly confirmed by [16]) and, therefore, the policy implications stemming from them. Yet the evidence supporting this choice is not ironclad. Understanding the properties of the tail is crucial to understand shock propagation in dynamic networks. The authors of Refs. [17, 18], among others, find that only a small fraction of possible network structures may spread relatively sizable contagion losses across the system, thus highlighting the non-linear nature of shock propagation effects and

stressing that contagion is to a considerable extent a tail risk problem.

This paper contributes to the debate by providing a detailed analysis of the network characteristics of a real interbank network over an extended period of time. We use daily data on bilateral interbank exposures and monthly bank balance sheets to study network characteristics of the Russian interbank market over Aug 1998 - Oct 2004. Among other things, the analysis allows one to determine the theoretical distributions of connectivity among banks via interbank loans, crucial to assess efficiency and stability of the Russian interbank market. We focus on measures that represent essential input for most of the interbank contagion simulations. Specifically, we examine the distributions of (un)directed (un)weighted degree, nodal attributes (bank assets, capital and capital-to-assets ratio) and edge weights (loan size and counterparty exposure). Using the methodology of [11] we set up a horse race between the different theoretical distributions to find one that fits the data best. We then study the time evolution of the best-fit parameters.

We observe that all studied distributions are heavy tailed. The fat tail typically contains 20% of the data and can be systematically described by a truncated power law. In most cases, however, separating the bulk and tail parts of the data is hard, so we proceed to study the full range of the events. We find that the stretched exponential and the log-normal distributions fit the full range of the data best. Our conclusions turn out to be robust to whether we aggregate the data over a week, month, quarter or year. Further, we find no qualitative difference between the "growth" and "maturity" phases of interbank market development and little difference between the "crisis" and "non-crisis" periods.

Sec. 2.3 describes our data, defines the network measures we study, and summarizes the conclusions from previous studies of those measures. Sec. 2.4 and 2.5, respectively, describe and illustrate the methodology we use to find the theoretical distribution that fits the data best. Sec. 2.6 reports the results. Sec. 2.7 concludes.

2.3 Data and Definitions

2.3.1 Data Source

Mobile, a private financial information agency, provided us with two datasets for the period Aug 1998 - Oct 2004 [1]. The information in the datasets is a part of standard disclosure requirements and is supplied to the regulator on a monthly basis. The first dataset, described in [19], contains monthly bank balances for most Russian banks. From this dataset we take two variables: total assets and capital. The second dataset contains monthly reports "On Interbank Loans and Deposits" (official form's code 0409501) and represents a register of all interbank loans issued in the Russian market. For each loan we know its size, interest rate,

[1] For more information on the data provider see its website at `www.mobile.ru`.

Country	Paper	Sample Period	Aggregation Level	# Banks
Mexican	[20]	2005 - 2010	day	40
Italy	[21]	1999 - 2002	day	$\pm$ 200
Italy	[22]	1999 - 2010	day/quarter	$\pm$ 200
Japan	[7]	2009m3 2013m3	month	125
Brazil	[4]	2007 - 2008	month	2400
Austria	[5]	2000 - 2003	month (only 10)	900
Germany	[23, 24]	2002q1 - 2012q3	quarter	2000+

Table 2.1: The levels of aggregation and the number of banks for earlier empirical studies of interbank markets (overnight as well as credit networks).

issuer, receiver, reporting date and maturity date. On average, about half of the Russian banks are active on the interbank market. Consequently, the analysis of interbank network measures includes fewer banks than the analysis of balance sheet indicators. With regard to the maturity of the loans, we discriminate between short-term and long-term loans. Short-term loans are defined as loans with a one-day or a one-week maturity. In this paper, we restrict ourselves to short-term loans which account for more than 80% of the transactions both in terms of the number and of the volume. The reasons for this restriction is because the data provide information about the repayment date and not about the issuance date of the loans. This makes it hard to infer the exact duration of the connection between two banks for the long-term loans.

2.3.2 Network Measures

Before we start constructing an interbank network from our data, we stress that there are several different ways of how banks can interact. For example, one can have liability [4, 5, 7, 20, 23, 24] and overnight lending networks [21, 22], but one can also construct networks of the financial payment flows between banks, or in aggregated form between countries [20, 25–29]. Banks are also connected by mutual cross holdings [30]. In this work we choose to consider all loans between banks with a maturity of less than a week. Hence, we construct an interbank network which combines the overnight market with the short-term liability market. This enables us to keep our finger on the pulse of the interbank market, by not including longer and hence stickier contracts.

We use the two aforementioned datasets to construct an interbank network with banks as nodes and mutual contracts representing edges. This procedure is performed for aggregated data covering various time intervals. In every situation, we construct three versions of the interbank network, which differ in the level of detail in quantifying the edges. In the undirected version, an undirected unweighted

edge is established between banks if they exchange money on the interbank market during the considered time period. For the directed versions, we discriminate between the issuer and the receiver of the interbank loan. A directed edge, which points from issuer to receiver, is established if the issuer has lent money to the receiver in the considered period. In the multidirected version, every interbank loan is represented by a directed edge pointing from issuer to receiver. This implies there can be several edges between pairs of nodes. The directed and undirected versions represent complementary views of the interbank network. The directed version captures the lending/borrowing activity by looking at the direction of the flows and therefore the contribution of each bank, whereas the undirected version merely captures the existence of interbank relations.

In this paper we focus on network characteristics that represent essential input for typical interbank contagion simulations [7, 15, 17, 18, 31–34]. We distinguish three types of network characteristics: nodal attributes, edge weights and various measures of a node's centrality.

We consider three nodal attributes: bank capital, total assets and leverage. Total assets proxy for bank size. Capital measures its financial buffer. Leverage is defined as capital divided by total assets. Upon evaluating nodal attributes, we exclude banks with negative assets (which we attribute to data errors) or negative capital (banks in effective default). In determining the distributions of edge weights and of node centrality measures we do not exclude banks with negative assets or negative capital. Edge weights are measured separately for the directed and the multidirected network. For the former, the edge weight equals counterparty exposure, that is, the total amount of money the issuer has lent to the receiver in the considered time period. For the multidirected network, the edge weight equals the size of the issued loan.

We consider various measures of a node's centrality. For an undirected network, we define the degree of a node as the number of edges connected to that node. It measures the number of counterparties a bank has on the interbank market. For the directed network, an in-degree (out-degree) of a node is the number of incoming (outgoing) edges; it measures the number of interbank lenders (borrowers) a bank has. For the multidirected network, an in-degree (out-degree) is the number of received (issued) loans. Finally, for each bank we define the total in-exposure (out-exposure) as the total amount of money borrowed from (lent to) the market.

The interbank credit network characteristics we consider in this paper have already been studied for a number of countries. Before sketching the main findings of those studies, we would like to draw the attention to Table 2.1 which summarizes the sheer variety of considered aggregation periods and of the actual networks' size. Because of this variation, care needs to be taken when comparing our results with those of other studies.

The majority of studies find numerous heavy tailed variables in the interbank network. In addition, many authors propose a power law as the best-fit candidate to (at least the tail of) the empirical distribution. In particular, Goddard et al. [8] find that the asset size distribution of U.S. commercial banks is well described by a truncated lognormal, while the tail part is well fitted with a power law. Cont et al. [4] obtain fair fits to the tails of the in-degree, out-degree, total degree, and exposure size distributions of the Brazilian interbank network with power laws. In the Austrian interbank market the loan size distribution is well described by a power law [5, 6]. That same study also examines the degree distribution of the undirected network and the in- and out-degree distribution of the directed alternative. Each of these three distributions is seemingly well described by two power laws, one for the low-degree region and one for the tail part. For Japan, Kanno et al. [7] find a power-law distribution for the entire range of the counter-party exposure size distribution. Finally, Iori et al. [21] investigate the distributions of the in-degree, out-degree and exposure of the individual banks in the Italian market. The authors do not attempt to fit their data with some theoretical distribution. They do find, however, a structural difference in the network topology between the months of the global financial crisis of 2008 and normal operating months.

Few studies, however, consider alternate theoretical distributions as candidates to describe their data. The ones that do [22], in fact, cast doubt on the ability of power laws to provide the best empirical fit. In what follows, we too consider an extensive list of alternative theoretical distributions and identify those that describe both the tail and the entire distribution better than the power law. Along the way, we apply a criterion which allows us to discriminate between the tail and the bulk parts of the data.

2.3.3 Descriptive Statistics

Different phases of the interbank market development (for example, growth versus maturity, or crisis versus non-crisis) may be guided by a different data generating mechanism. For example, Iori et al. [21] find a structural difference in the Italian interbank network topology before versus during the 2008 financial crisis. This finding suggests that the choice of the distribution that best fits the data may vary over time, and in particular may differ between crisis and non-crisis periods. In Sect. 2.6 we check whether such variation exists. In this section, we provide evidence that the Russian interbank market, indeed, went through some distinct development phases.

Fig. 2.1 shows the time series of the number of banks active on the interbank market and of the number of issued loans. For this figure we use data aggregated over a month. Over 1998-2002 the interbank network experiences growth: we observe a steady and comparable increase in the number of active banks and of

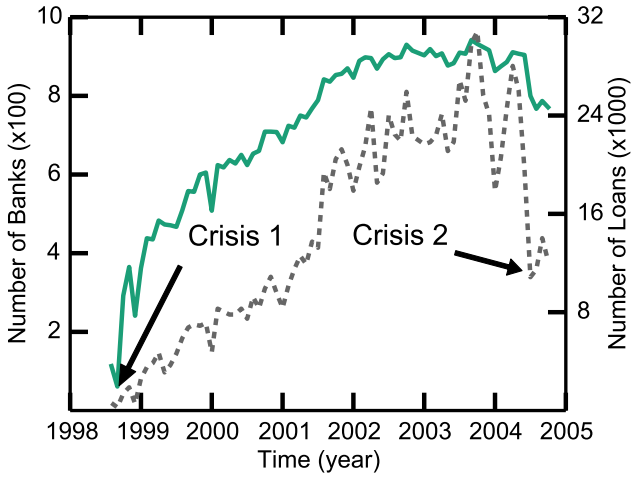


Figure 2.1: (Color online) The time dependence of the number of active banks (solid line) and the number of interbank loans (dashed line) in the Russian interbank network between 1998 and 2005. Data are aggregated over a month. The arrows indicate the start of the two "crises".

issued loans. After 2002, the interbank network gradually matures: the number of active banks flattens out while the number of loans per bank grows. Note, however, the strong variation in the number of issued loans from the second half of 2003 onwards.

As is clearly indicated in Fig. 2.1, our sample period includes two crises: one in August 1998 and one in the summer of 2004. Both crises resulted in a partial meltdown of the Russian interbank market. They coincide with the edges of the sample period and are clearly marked by a reduction in the number of active banks and issued loans. The first crisis got triggered on August 17, 1998 when Russia abandoned its exchange rate regime, defaulted on its domestic public debt and declared a moratorium on all private foreign liabilities. The second crisis was ignited by an investigation of banks accused of money laundering and sponsorship of terrorism. This gave rise to a wave of distrust among banks and a consequent liquidity drought.

Fig. 2.2 illustrates the impact of Aug 1998 crisis on the Russian interbank network. Nodes are banks, and directed edges represent issued loans. The first panel shows the activity in the market in the two weeks leading up to the seizure, whereas the second panel covers the two weeks after the collapse. Evidently, we see a decrease in the number of active banks, from 87 to 65, and in the number of loans granted, from 507 to 96. When considering the structure in this unweighted

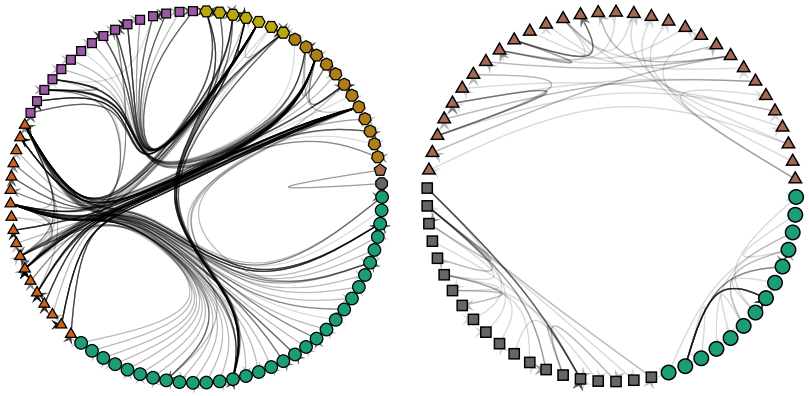


Figure 2.2: (Color online) The activity on the interbank market in the two weeks leading up to August 17, 1998 (left panel) and the two consecutive weeks (right panel). A bank is represented as node and each link represents an issued loan. Nodes are colored per block as defined in Ref. [35].

multidirected network we can find a clear distinction between the "normal" and "crisis" periods. The technique of Ref. [35] can be used to uncover groups of nodes, called blocks, which fulfil similar roles in the network [2]. In the two weeks leading up to the crisis we get an interbank market with 5 blocks (and two isolated banks). After the crisis the number of identified blocks goes down to 3. Notably, the identified blocks interact with each other before the crisis but not after. So the crisis disintegrates the network and reconfigures banks' role in it.

We can also see evidence of different development phases by studying the time evolution of two basic network measures: the average local clustering coefficient, which is a potential indicator of systemic risk [36], and the average shortest path length. For an undirected network, the local clustering coefficient $0 \leq c_u \leq 1$ of a node u is the number of the edges between the nodes of the neighborhood of u divided by the maximum amount that could possibly exist between them. It can be conveniently defined as in [37]

$$c_u = \frac{2T(u)}{deg(u)\,(deg(u)-1)}, \tag{2.1}$$

where $T(u)$ is the number of triangles (a subgraph with three nodes and three edges) attached to u and $deg(u)$ is the degree of u. The average of this local

[2]The technique of Ref. [35] makes use of a hierarchical version of stochastic block modeling and offers an alternative to the usual modularity optimization algorithms.

clustering coefficient is

$$C = \frac{1}{n} \sum_{u \in G} c_u, \tag{2.2}$$

where n is the number of nodes in the network G [38]. The average shortest path length D is defined as

$$D = \sum_{u,v \in V} \frac{d(u,v)}{n(n-1)} \tag{2.3}$$

where V is the set of nodes in the network G, and $d(u,v)$ is the length of the shortest path from node u to v [39]. In a disconnected graph D can only be computed for the largest connected component.

Fig. 2.3 shows the time evolution of C and D computed per month for the largest connected component of the undirected interbank network. First, note that the local clustering coefficient averaged over all nodes and time periods ($C = 0.198$) is nearly identical to the one reported by [4] for the Brazilian interbank market ($C = 0.2$) but is much lower than for the German market (directed clustering $C = 0.80$) [23]. In contrast, the average shortest path length ($D \approx 3$) is notably higher compared to German ($D = 2.14$) [23], to Austrian ($D = 2$) [5] and Mexican ($D = 1.7$) [20] interbank networks. In the case of the Mexican and Austrian ones, this difference is likely driven by the fact that the Russian network is sizably larger.

The C and D tend to move in the opposite direction: three months after the first crisis hits C drops while D spikes; during the growing phase of the network (1999-2002) C grows while D falls; during the mature phase (2002-2004) both measures stabilize at $C \approx 0.22$ and $D \approx 3$; finally, during the 2004 crisis C drops while D spikes again. The average local clustering coefficient C, however, tends to have bigger fluctuations from period to period. In contrast, the time series of the average shortest path length is very smooth, and the only two obvious spikes occur around the two crises. Clearly, those crises disrupted the overall network structure: the liquidity drought, which is equivalent to the pruning of links, caused a significant increase in the average shortest path length. Even the decrease in the number of nodes, and hence shrinking of the network, could not offset this effect. The 1998 spike in D is particularly remarkable given we only consider the largest connected component, that is, about one third of the nodes (see Fig. 2.2).

2.4 Methodology

For each network measure we test whether the power law or an alternate fat-tailed distribution fits the data best. First, we fit distributions to the tail, then to the entire data range. Along the way, we apply a criterion which allows us to discriminate between the tail and the bulk parts of the data.

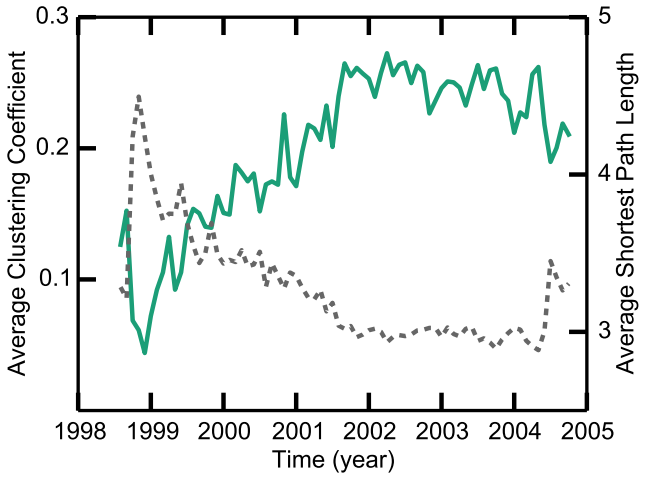


Figure 2.3: (Color online) The time dependence of the average clustering coefficient (solid line) and the average shortest path length (dashed line) calculated for the most connected component of the undirected version of the Russian interbank network between 1998 and 2005. Data are aggregated over a month.

Distribution	$f(x)$
Power law (PL)	$Cx^{-\alpha}$
Truncated power law (TPL)	$Cx^{-\alpha}e^{-\lambda x}$
Exponential (Exp)	$Ce^{-\lambda x}$
Stretched exponential (SExp)	$Cx^{\beta-1}e^{-(\lambda x)^{\beta}}$
Log-normal (LN)	$\frac{C}{x}\exp\left[-\frac{(\ln x-\mu)^2}{2\sigma^2}\right]$

Table 2.2: Definition of the normalized distributions used in this work. The constant C is defined by the normalization condition $\int_{x_{min}}^{\infty} f(x)dx = 1$.

To study the tail we adopt the methodology of Ref. [11]. First, we fit a power law (PL) to the data to determine the starting point of the tail part (the so-called cut-off x_{min} of the scaling range). Then we fit each of the theoretical distributions to the tail using maximum likelihood (ML). Finally, we use a relative goodness-of-fit test to select the distribution that fits the data best. We consider the same selection of theoretical distributions as in Ref. [11]. Their functional forms are listed in Table 2.2 and involve one or two free parameters.

For completeness and to introduce the notation we briefly sketch the methodology of Ref. [11]. Consider a given ordered data set $\{x_j, j = 1, \ldots, N\}$. Every entry x_j is a potential x_{min} and for each of those we compute the ML estimate of

the power-law exponent α

$$\hat{\alpha}\left(x_j = x_{min}\right) = 1 + (N - j + 1)\left[\sum_{i=j}^{N} \ln \frac{x_i}{x_{min}}\right] . \tag{2.4}$$

We then use the Kolmogorov-Smirnov test to select the optimum $\hat{x}_{min}$. It is defined as the cut-off which minimizes the quantity

$$Z = \max_{x \geq x_{min}} \mid S(x) - P(x) \mid . \tag{2.5}$$

Here, $S(x)$ is the cumulative distribution function (CDF) of the observed values for $x_j \geq x_{min}$, and $P(x) = \frac{C}{-\alpha+1}\left(x^{-\alpha+1} - x_{min}^{-\alpha+1}\right)$ is the CDF of the power-law fit to the tail part of the data.

In the next step, for given x_{min}, we perform ML fits to the data with the other four candidate distributions of Table 2.2. As argued by Ref. [11], it is more useful to know which distribution is the best possible fit candidate, rather than the goodness of fit for each distribution individually. To compare the relative goodness of the different fits, we compute the likelihood ratios R for the pairs of probability density functions (PDFs) $p_1(x_i)$ and $p_2(x_i)$

$$R\left(p_1, p_2\right) = \frac{L_1}{L_2} = \prod_{i=j}^{N} \frac{p_1(x_i)}{p_2(x_i)} . \tag{2.6}$$

The corresponding normalized loglikelihood ratios $\mathcal{R}\left(p_1, p_2\right)$ read

$$\begin{aligned} \mathcal{R}\left(p_1, p_2\right) &= \frac{1}{\sigma_{12}\sqrt{N - j + 1}} \sum_{i=j}^{N} \left[\ln \frac{p_1(x_i)}{p_2(x_i)}\right] \\ &= \frac{1}{\sigma_{12}\sqrt{N - j + 1}} \sum_{i=j}^{N} \left[l_i^{(1)} - l_i^{(2)}\right] , \end{aligned} \tag{2.7}$$

where $l_i^{(k)} = \ln p_k(x_i)$ and σ_{12} is defined as

$$\sigma_{12}^2 = \frac{1}{N - j + 1} \sum_{i=j}^{N} \left[\left(l_i^{(1)} - l_i^{(2)}\right) - \left(\bar{l}^{(1)} - \bar{l}^{(2)}\right)\right]^2 . \tag{2.8}$$

The $\mathcal{R}\left(p_1, p_2\right)$ is positive (negative) if the data is more likely in the p_1 (p_2) distribution.

In order to guarantee that the value of $\mathcal{R}$ is not merely a product of fluctuations and that the true expectation value of $\mathcal{R}$ is zero, we compute the probability that the measured normalized log likelihood ratio has a magnitude as large or larger than the observed value $|\mathcal{R}|$. This so-called p-value is defined as

$$p\left(\mathcal{R}\right) = \frac{1}{\sqrt{2\pi}}\left[\int_{-\infty}^{-|\mathcal{R}|} e^{-\frac{t^2}{2}} dt + \int_{|\mathcal{R}|}^{\infty} e^{-\frac{t^2}{2}} dt\right] . \tag{2.9}$$

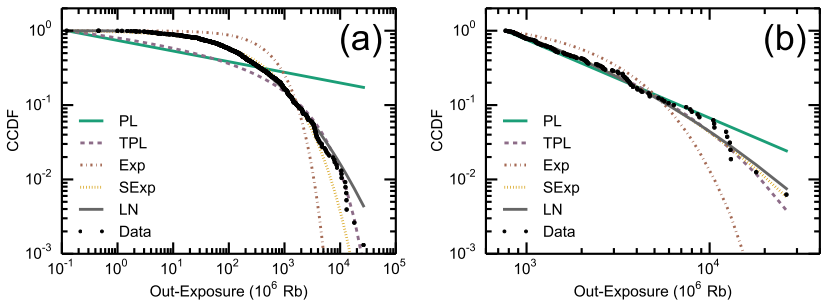


Figure 2.4: (Color online) The complementary cumulative distribution functiom (CCDF) of the entire (a) and tail part (b) of the monthly-aggregated out-exposure distribution for January 2002. Also shown are the ML fits to the CCDFs for the five distributons of Table 2.2.

The distribution p_k with the highest value of

$$g\left(p_k\right) = \sum_{l \neq k} \mathcal{R}\left(p_k, p_l\right) \ , \tag{2.10}$$

is considered as the most suitable distribution among the different candidates $\{p_i\}$. Bootstrapping and the Kolmogorov-Smirnov test are alternate methods to compute the p values. Both methods, however, are subject to some pitfalls as outlined in Ref. [40]. By using the PL as a benchmark and letting it decide on the value of x_{min}, we are confident that our methodology produces the best possible PL fit to the data. If it turns out that an alternate distribution provides a significantly better fit to the tail part of the data than the PL, strong evidence emerges that this theoretical distribution better accounts for reality than a power law.

We stress that the above-sketched methodology of Ref. [11] can not only be applied to the analysis of the tail part of the data but also to the full range of the data by setting $x_{min} = x_1$. When considering the full range of data it is natural to select a set of theoretical distributions that encompass a Gaussian-like regime (thermal or bulk part) supplemented with a fat tail (superthermal part). Examples of such distributions include the stretched exponential and the log-normal.

2.5 Illustration of Methodology for Out-exposure distribution

As a prototypical example of the methodology sketched in Sec. 2.4, we explore the out-exposure distribution for the monthly aggregated data. The data for the out-exposure distribution of January 2002 shown in Fig. 2.4 are exemplary for all

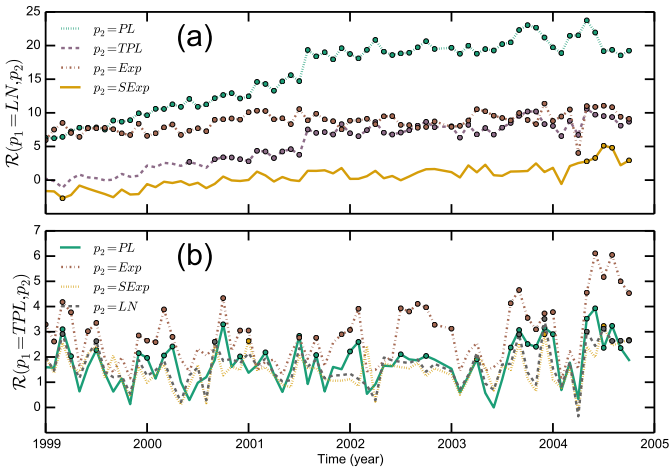


Figure 2.5: (Color online) (a) The time series of the normalized loglikelihood ratios $\mathcal{R}(p_1 = LN, p_2)$ of Eq. (2.7) for the different fits to the total distribution of the monthly-aggregated out-exposure. (b) The time series of $\mathcal{R}(p_1 = TPL, p_2)$ for the fit to the tail of the monthly-aggregated out-exposure distribution.

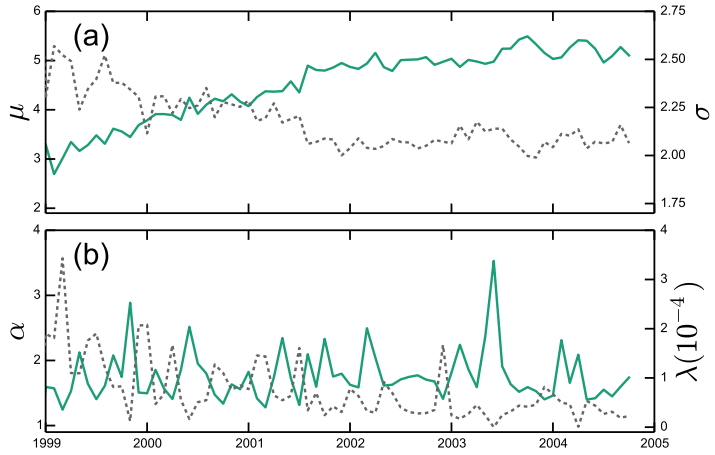


Figure 2.6: (Color online) The time evolution of the parameters for the best fit to the total (a) and tail (b) part of the out-exposure distributions. The total data range is fitted with a log-normal distribution with parameters μ (solid line) and σ (dashed line). The tail part is fitted with a truncated power-law distribution with parameters α (solid line) and λ (dashed line).

Table 2.3: Summary of the results of the fits to the tails of the monthly-aggregated data of all the network features considered in this work. We list the average fraction of the events in the tail and the distribution which provides the best fit, together with the percentage of months for which it comes out "best" according to the criterion of Eq. (2.10). Columns 4 and 5 provide the average values of the parameters of the best-fit distributions. The last column lists the distributions which are not significantly (1 percent level) worse for more than half of the 70 months (January 1999 - October 2004) considered, together with the percentage of months for which this is the situation. The measures with an extra label "(B)" refer to an analysis of a sample obtained after bootstrapping the real data.

Measure	Tail (%)	Best fit (%)	α	λ	Alternate fits (%)
Asset Size	23 ± 9	TPL (77)	1.80 ± 0.06	$(4.47 \pm 9.56) \times 10^{-7}$	PL (100), SExp (100), LN (97)
Capital Size	31 ± 18	TPL (58)	1.91 ± 0.14	$(1.25 \pm 1.71) \times 10^{-5}$	PL (100), SExp (100), LN (89)
Leverage	24 ± 9	PL (57)	3.67 ± 0.76	None	TPL (100), Exp (86), SExp (100), LN(100)
Leverage (B)	28 ± 10	TPL (57)	3.77 ± 0.10	$(1.81 \pm 2.77) \times 10^{-6}$	PL (100), Exp (93), SExp (100), LN(100)
Loan Size	5 ± 9	TPL (75)	2.49 ± 0.47	$(3.22 \pm 3.54) \times 10^{-4}$	PL (91), SExp (87), LN (71)
Counterparty-Exposure	12 ± 8	TPL (80)	2.17 ± 0.51	$(2.53 \pm 2.44) \times 10^{-4}$	PL (53), SExp (81), LN (71)
Counterparty-Exposure (B)	16 ± 9	TPL (86)	2.02 ± 0.36	$(3.13 \pm 2.33) \times 10^{-4}$	PL (91), SExp (100), LN (100)
Undirected Degree	21 ± 16	TPL (100)	1.80 ± 0.36	$(1.83 \pm 0.83) \times 10^{-2}$	PL (70), Exp (68), SExp (97), LN (97)
Directed In-Degree	25 ± 20	TPL (94)	1.72 ± 0.43	$(2.65 \pm 2.74) \times 10^{-2}$	PL (71), Exp (68), SExp (91), LN (95)
Directed Out-Degree	13 ± 8	TPL (89)	2.07 ± 0.66	$(2.93 \pm 1.92) \times 10^{-2}$	PL (85), Exp (98), SExp (100), LN (100)
Directed Out-Degree (B)	28 ± 18	TPL (77)	1.94 ± 0.71	$(4.13 \pm 3.99) \times 10^{-2}$	PL (99), Exp (100), SExp (99), LN (99)
Multidirected In-Degree	23 ± 24	TPL (97)	1.83 ± 0.48	$(5.05 \pm 3.93) \times 10^{-3}$	PL (74), Exp (77), SExp (99), LN (100)
Multidirected Out-Degree	23 ± 17	TPL (93)	1.74 ± 0.50	$(7.80 \pm 5.83) \times 10^{-3}$	PL (63), Exp (88), SExp (93), LN (91)
In-Exposure	30 ± 17	TPL (97)	1.49 ± 0.43	$(7.68 \pm 5.11) \times 10^{-5}$	SExp (57)
Out-Exposure	20 ± 11	TPL (94)	1.76 ± 0.39	$(7.49 \pm 6.23) \times 10^{-5}$	PL (59), SExp (94), LN (83)

considered 70 months (January 1999 - October 2004). The considered values for the out exposure extend over more than five orders of magnitude and the distribution can be labelled as heavy-tailed [3].

Using the methodology outlined in Sec. 2.4 we identify the tail part of the data. From Fig. 2.4(b) it is clear that fits to the tail part of the considered data with the two-parameter log-normal (LN), stretched exponential (SExp) and truncated power law (TPL) distributions outperform those with the one-parameter exponential (Exp) and power law (PL) distributions. The parameter λ in the TPL accounts for the finite-size effects near the upper edge of the out-exposure distribution. For each of the 70 months in our sample we compute the ratios $\mathcal{R}\left(p_1, p_2\right)$ of Eq. (2.7) for all pair combinations out of the list of five distributions of Table 2.2. In Fig. 2.5 we display the time series of the $\mathcal{R}\left(\text{TPL}, p_2\right)$ for p_2 =PL, Exp, SExp, and LN. We observe that $\mathcal{R}\left(\text{TPL}, p_2\right)$ is mostly positive, which indicates that the TPL offers the best overall description of the tail of the out-exposure data. In order to test the significance of this observation, we evaluate the p values. If the normalized ratio for a pair of distributions in a given month is significant at a one percent level, the data points in Fig. 2.5 are dotted. The figure indicates that the two-parameter TPL is a significantly better fit to the tail of the monthly-aggregated out-exposure distribution than a power law and an exponential. The TPL, however, does not provide a significantly better fit than the LN or SExp for most of the months.

Fig. 2.6(b) shows the time evolution of the TPL parameters. Whereas α fluctuates around the same average value, λ falls over time. Hence, as time proceeds, the exponential cutoff to the power law shifts to larger values of the out-exposure.

To weigh the relative performance of the different distributions we compute their monthly $g\left(p_i\right)$ scores (see Eq. (2.10)). We dub p_i as the "best" overall fit candidate when its $g\left(p_i\right)$ score is highest for the largest fraction of the 70 months in our sample. As can be seen in Table 2.3, the truncated power law is the best-fit candidate in 94% of the considered months. Nevertheless, $\mathcal{R}\left(\text{TPL}, \text{SExp}\right)$ is not significantly different from zero in 94% of the considered months as is the case with $\mathcal{R}\left(\text{TPL}, \text{PL}\right)$ in 59% and with $\mathcal{R}\left(\text{TPL}, \text{LN}\right)$ in 83% of the months. So we list the SExp as well as PL and LN as alternate best-fit candidates. In general, if a distribution is not significantly worse than the best-fit candidate in more than half of the months, it is mentioned in the last column of Table 2.3. We conclude that although the tail is described best by a TPL, the fit is not significantly better than the power-law, stretched-exponential and log-normal fits.

We now turn our attention to the distributions covering all out-exposure data points, from small to large. Again we consider the five PDFs of Table 2.2. From the ML fits, for example for January 2002 displayed in Fig. 2.4(a), we immediately notice that the exponential and power law do not fit the entire range of data well. This is a clear indication of the fact that the distribution of the out-exposure has

[3] A distribution is defined as heavy-tailed if it is not exponentially bounded [40].

both a thermal (Gaussian-like) and a superthermal (fat tail) part. Fig. 2.5 shows the normalized loglikelihood ratios $\mathcal{R}(p_1 = \text{LN}, p_2)$ over time. Using the same criteria as for the tail part, we conclude that the log-normal is the best fit candidate for 80% of the 70 months studied.

We also find that the stretched exponential is not significantly worse in 57% of the cases, which is reflected in the persistently small values of their likelihood ratios in the top panel of Fig. 2.5. During the 2004 crisis the stretched exponential has a significantly better fit than the log-normal. As the number of interbank loans is subject to a sudden drop in the summer of 2004 (Fig. 2.1), this change in the best-fit candidate illustrates how the network adapts to changing overall conditions. Fig. 2.6(a) shows the time evolution of the log-normal fit parameters. The parameter μ increases during the "growth" stage of the network whereas σ decreases, and both tend to flatten out in the "mature" phase of the network when few nodes or links are added (see Fig. 2.1).

Now, we investigate in how far the size of the time bin widths affects the conclusions with regard to the "best" theoretical distributions describing the data. To this end, we study the out-exposure and consider different time intervals to aggregate the data. Beside the networks build with all loans issued in one month, we can do this for all loans issued in a week, quarter or year. Fig. 2.7 shows the out-exposure distribution for the second week of January 2002, January 2002, the first quarter of 2002, and the entire year of 2002. We also include the ML fit with the stretched exponential and log-normal for each of the aggregates. If we calculate $g(p_k)$ for every distribution in Table 2.2, we find that the log-normal is the preferred candidate to fit each dataset, although, again, the stretched exponential is not significantly worse. For the tail part, the truncated power law stays on top too. Because the data has a monthly periodicity, linked to the monthly compliance with regulatory requirements, aggregating the data monthly is the most natural thing to do.

To check whether these observations can be generalized, we repeat the illustrated procedure for weekly, quarterly, and yearly aggregates. Tab. 2.5 shows the "best" fit candidate for the tail and entire distribution over our data sample. For the tail part, the TPL is preferred over the others for each aggregate. It does so in 85% of the weeks and up to 100% of the years. For the entire distribution, the LN is always preferred.

2.6 Results

The methodology used for analysing the out-exposures in the previous section, is used for each of the 12 complementary interbank network measures introduced in Sec. 2.3 and listed in Table 2.3. For a few combinations of network variables and time instances, we find that the x_{min}, which marks the lower boundary of

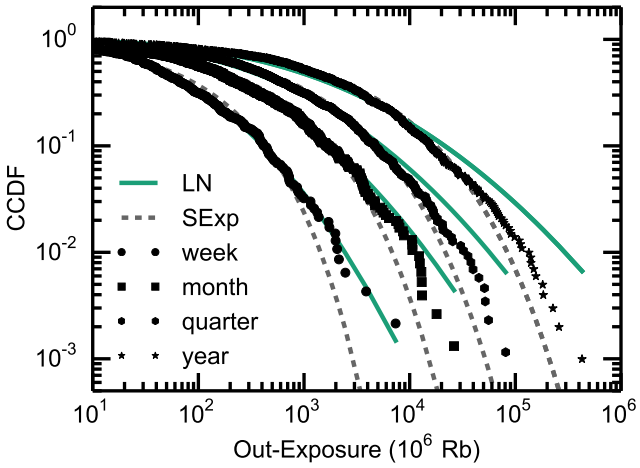


Figure 2.7: (Color online) The CCDF of the out-exposure distribution from a network aggregated with data from one week (week 2 of January 2002), from one month (January 2002), from one quarter (first quarter of 2002), and one year (2002). We also show the respective stretched exponential and log-normal fits.

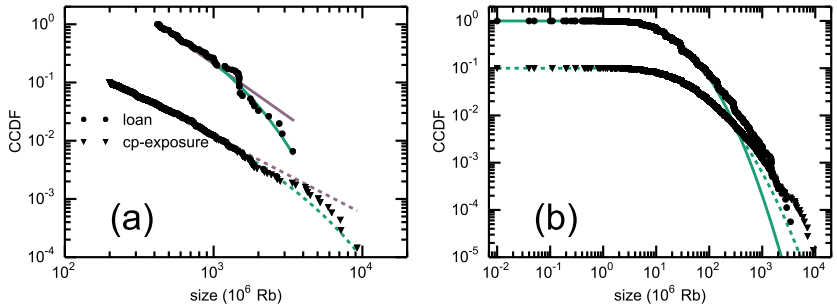


Figure 2.8: (Color online) The CCDF for the loan-size and the counterparty(cp)-exposure for January 2004. Panel (a) shows the tail part of the data together with a PL (grey) and TPL (green) fit. Panel (b) shows the data over the entire range together with LN fit. For illustrative reasons, the data for the cp-exposure have been scaled by a factor of 0.1.

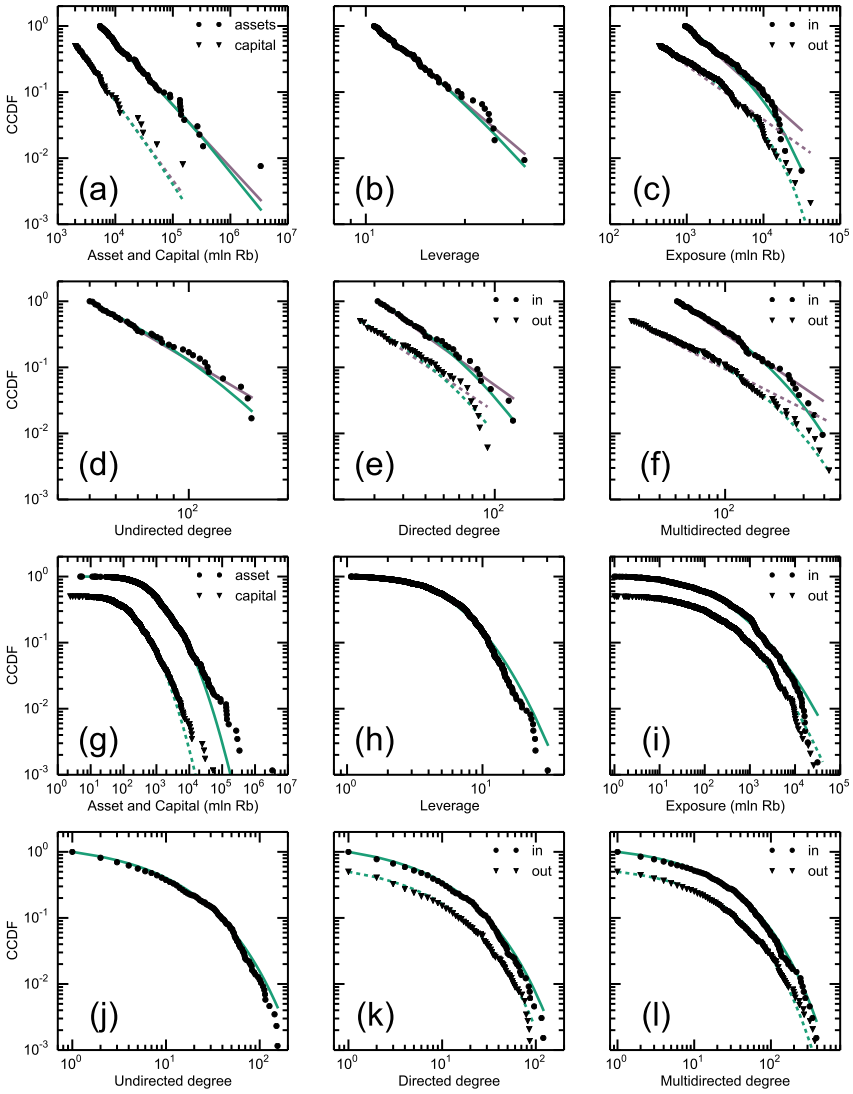


Figure 2.9: (Color online) The CCDF for 10 *interbank network measures for the montly aggregated data in January 2004. Panels (a-f) display the tails parts of the data together with the PL (grey) and TPL (green) fits. Panels (g-l) contain the data over the entire range together with the "best" fit from Table 2.4. For illustrative reasons, in those panels with two measures, the data of the second measure have been shifted vertically by multiplying them by a scaling factor of* 0.5.

the tail parts of the distributions and results from minimizing the quantity Z of Eq. (2.5), is close to the upper edge of the distribution. Under those circumstances the identified tail does not hold a sufficient amount of datapoints to perform a meaningful fit. This is particularly common during the 1998 interbank network collapse. For this reason, we only study the best-fit parameters from January 1999 onwards.

A prototypical example of a data set for which the fit with the functions of Table 2.2 is not fully satisfactory is shown in Fig. 2.8. The figure includes the loan size and counterparty-exposure measures for January 2004. Whereas the TPL is a good match for the tail parts of the data, the "best" fit to the entire range of the data with a LN distribution clearly underestimates the probablility of events in the tail. Fig. 2.9 displays for January 2004 the "tail" and "bulk+tail" parts of the 10 other interbank network measures considered. It is clear that the data can be reasonably well described by the adopted procedure and proposed theoretical distributions.

Table 2.3 reports our findings on the tail parts of the empirical distributions. The second column shows the percentage of data points assigned to the tail via the Kolmogorov-Smirnov criterion of Eq. (2.5). This number averages to about 20% across the considered network measures but varies a lot. In particular, the loan size, counterparty exposure, and directed out-degree have fewer than 20% of the data-points in the tail. For most of the 70 months included in the analysis, the truncated power law outperforms, according to our methodology, the other four candidates for every measure. In Table 2.3 the time-averaged values for the α and λ parameters are also reported. The power α is consistently of the order of two. It emerges from our studies that a fat tail is a characteristic feature of the measurable quantities of an interbank system. We notice that λ is very volatile.

Upon evaluating the last column of Table 2.3, we find that although the truncated power law systematically provides the best fit, the other candidates are hardly significantly worse at the 1 percent level. This observation is particularly true for the unweighted degree measures which are discrete. In fact, for the degree measures all directed and undirected versions lead to similar results. We find that none of the five distribution candidates can be conclusively labeled as representing the best fit to the degree data. We note that the fraction of events located in the tail fluctuates a lot.

In order to evaluate the degree of sensitivity of our results to the amount of observations in the tail, we have made use of bootstrapping. For some prototypical measures we have drawn 1000 bootstrapped samples (random samples with replacement) of the real observations. The results of the analysis of the bootstrapped sample for the leverage, counterparty-exposure and directed out-degree are also included in Table 2.3. For the directed out-degree and the counterparty exposure, the fits to the real and bootstrapped data provide results which are compatible within the error bars. The leverage is the only studied network measure for which the best

fit to the real data is provided by a PL. The PL, however, is the best fit for only 57% of the studied months. For all studied months, the TPL emerges as a good alternate fit to the leverage data. Accordingly, it is not all that surprising that a TPL emerges as the best fit to the bootstrapped leverage data. All by all, we may conclude that despite the fact that the number of observations in the tails is rather volatile, the fits to the tails are rather robust.

The fact that the fraction of events in the tail is very volatile hints that a more integrated approach whereby the bulk and the tail parts of the data are simultaneously fitted with a non-Gaussian distribution may lead to a more robust description of the network features.

Table 2.4 summarizes our fits to the entire distributions of all 12 network measures over all 70 months using monthly-aggregated data. The asset, capital, leverage, loan sizes, as well as the counterparty-exposure and in-exposure distributions are described significantly best by a log-normal. The various degree distributions prefer the stretched exponential, which is on average not significantly better than the log-normal and/or the truncated power law. Similar results are obtained for the directed and undirected versions of the degree distributions. For those network measures for which the stretched exponential represents the best fit, we observe that the parameter λ is very volatile. It is worth noting that the stretched exponential has been put forward as a natural fat tail distribution for many physical and economic phenomena which cannot be satisfactorily described by a power law [41]. In general, the parameters entering the fits to the tail parts are subject to larger fluctuations than those entering the distributions for the complete data set.

Our results compare well with some of the existing studies of interbank networks, but differ from others. Specifically, our findings are in line with Goddard et al. [8] who study U.S. banks: the log-normal distribution fits the bulk part of the asset-size data well, while the power law does the same for the tail part. Our findings are not too different from those of Cont et al. [4] for the Brazilian interbank network: while our TPL fits to the tails of various degree distributions (see Table 2.3) deliver values of α between 1.7 and 2.2, Cont et al. [4] report power-law exponents between 2.2 and 2.8 ($\alpha = 2.54$ for total degree, $\alpha = 2.46$ for in-degree, $\alpha = 2.83$ for out-degree, and $\alpha = 2.27$ for out-exposure size). Finally, in contrast to [5] studying the Austrian interbank market we do not find that the power law provides a good fit to the entire loan-size distribution ([5] report $\alpha = 1.87$). Their power-law exponents reported for the tails of total and in-degree distributions (resp. $\alpha = 2.0$ and $\alpha = 3.11$) are comparable to the TPL values in Table 2.3, while their out-degree $\alpha = 1.72$ differs substantially.

The most extensive study of interbank distributions was performed for the e-mid market [22]. In line with this work, the authors consider both the complete and tail parts of the distributions and report results for daily and quarterly time aggregates. For the quarterly aggregates of the in-, out- and total degree distribu-

tion, a stretched exponential emerged as the best fit to the entire distribution. This confirms our findings. A log normal was reported as a best fit to the tail parts of the data. We also find that the LN can reasonably account for the tail parts. It is slightly outperformed by the truncated power law (which was not included in the set of possible distributions in the work of Ref. [22]) and flags as not significantly worse in our table. For the quarterly aggregated number of transactions, a network feature comparable to the multidirected degree considered here, they also find similar results as for the regular degrees.

Upon scrutiny of the time evolution of the fitted distributions and their parameters, we do not find any systematic changes in the distributions which emerge as best fit to the data, between the "growth" phase (1999-2002) and the "mature" phase (2002-2004). In line with the expectations, however, the extracted parameters are subject to smaller variations as the network grows and the number of nodes and edges increases.

As already mentioned, our data set covers two crises, the first in August 1998 and the second in the summer of 2004. We would like to know if the best-fit candidates for the crisis periods are different than the ones for the normal operation periods.

As a consequence of the moratorium in August 1998, the interbank market collapsed in September, and then again in December 1998 (see Fig. 2.1). Due to this collapse, the network is too small and their are too few transactions to gather proper statistics. Fits to the data in this period are inconclusive for most of the measures, and hence are not included in the overall discussion.

During the second crisis period, the network does stay large enough to have proper statistics. From Fig. 2.5 it is clear that the stretched exponential distribution is significantly better than the log-normal from mid 2004 onwards. Thus, for the out-exposure (as well as the in-exposure), there is a difference between pre- and post-crisis best-fit candidates. For the other measures, however, we find that the preferred distributions are identical in "crisis" and in "normal" operation periods.

To end this section, we test the qualitative robustness of the obtained results using different time windows. This is particulary important in view of the fact that a recent study has shown that the interbank network properties depend on the aggregation period [42]. We stress that on average the number of banks active in one week is 77% of the monthly active ones. Accordingly, some banks are active in the interbank credit market during one week of a particular month. For the quarters and years this is respectively 114% and 132%. This increase is mainly driven by new entrants to the market. Hence, banks tend to participate in the market at least once per month. To check if the preferred distribution is the same for different time aggregates, we repeat the monthly procedure for weekly, quarterly, and yearly aggregates. Table 2.5 shows the "best" fit for each measure-aggregate pair and the percentage of the bins it is considered so. For the tail parts, we find that the TPL

Table 2.4: Summary of the results of the fits to the monthly-aggregated data of all the network features considered in this work. We list the distribution which provides the best fit to the full range of the data ("bulk"+"tail")), together with the percentage of months for which it comes out "best" according to the criterion of Eq. (2.10). Columns 3 and 4 provide the average values of the parameters of the best-fit distribution. The last column list the distributions which are not significantly (at the 1 percent level) worse for more than half of the 70 months (January 1999 - October 2004) considered, together with the percentage of months for which this is the situation.

Measure	Best fit (%)	Parameter 1	Parameter 2	Alternative fit (%)
Asset Size	LN (100)	$\mu = 5.45 \pm 0.63$	$\sigma = 1.91 \pm 0.02$	None
Capital Size	LN (100)	$\mu = 4.64 \pm 0.53$	$\sigma = 1.51 \pm 0.03$	None
Leverage	LN (99)	$\mu = 1.57 \pm 0.05$	$\sigma = 0.74 \pm 0.01$	None
Loan Size	LN (100)	$\mu = 2.54 \pm 0.34$	$\sigma = 1.27 \pm 0.08$	None
Counterparty-exposure	LN (99)	$\mu = 3.33 \pm 0.31$	$\sigma = 1.56 \pm 0.13$	None
Undirected Degree	SExp (54)	$\lambda = 0.29 \pm 0.20$	$\beta = 0.55 \pm 0.06$	LN (94)
Directed In-Degree	SExp (66)	$\lambda = 0.41 \pm 0.27$	$\beta = 0.46 \pm 0.07$	TPL (91), LN (71)
Directed Out-Degree	SExp (83)	$\lambda = 0.43 \pm 0.39$	$\beta = 0.60 \pm 0.05$	TPL (51), LN (97)
Multidirected In-Degree	SExp (80)	$\lambda = 0.29 \pm 0.35$	$\beta = 0.43 \pm 0.07$	TPL (90)
Multidirected Out-Degree	SExp (73)	$\lambda = 0.19 \pm 0.26$	$\beta = 0.53 \pm 0.09$	LN (80)
In-Exposure	LN (83)	$\mu = 4.26 \pm 0.83$	$\sigma = 2.37 \pm 0.09$	None
Out-Exposure	LN (80)	$\mu = 4.55 \pm 0.65$	$\sigma = 2.13 \pm 0.10$	SExp (57)

can be considered as the best overall fit candidate. We notice that in general the TPL becomes more preferred as the bin width increases. For each measure, the favoured fits to the entire distribution stay the same for different bin widths. We emphasize that the stability of the best-fit distribution with respect to variations in time do not imply that the parameters are not subject to time variations.

2.7 Conclusion

In this paper we use daily data on bilateral interbank exposures and monthly bank balance sheets to study network characteristics of the Russian interbank market over Aug 1998 - Oct 2004. Specifically, we examine the distributions of (un)directed (un)weighted degree, nodal attributes (bank assets, capital and capital-to-assets ratio) and edge weights (loan size and counterparty exposure). Using the methodology of [11] we set up a horse race between the different theoretical distributions to find one that fits the data best.

In line with the existing literature, we observe that all studied distributions are heavy tailed with the tail typically containing 20% of the data. The tail is best described by a truncated power law, although the fit of other candidate distributions is only marginally worse. In most cases, separating the bulk and tail parts of the data turns out to be hard, and the proportion of observations assigned to the tail varies a lot. More stable fits to the data are obtained in an integrated approach that accounts for both the Gaussian and the non-Gaussian parts of the distributions. We find two distributions that fit the full range of the data best: the stretched exponential for measures related to unweighted degree and the log-normal for everything else. In case of the former, the log-normal performs only marginally worse. The power law distribution is rather ill suited to represent the full range of the studied characteristics of the Russian interbank market.

Our conclusions with regard to the best-fit distributions turn out to be robust to whether we aggregate the data over a week, month, quarter or year. Further, we find no qualitative difference between the "growth" and "maturity" phases of interbank market development and little difference between the "crisis" and "non-crisis" periods.

Our findings support the recent call [11, 12] for more rigorous statistical tests to detect power-law behavior in empirical data. While for most variables we find that the power law fits the tail of the distribution reasonably well, it is:

1. *almost never the best* candidate to describe the tail;

2. *typically not a good* candidate to describe the whole distribution.

Our findings echo those of Ref. [22] who also find that the power law provides an inferior fit, compared to alternative distributions, to their overnight money market

Table 2.5: Robustness of the "best" fits for different aggregates. This table shows for each measure-aggregate pair the preferred fit candidate as well as the percentage of the bins it was considered so. This is done for fit of tail as well as total.

	Tail				Bulk+Tail			
Measure	Week	Month	Quarter	Year	Week	Month	Quarter	Year
Loan Size	TPL (70)	TPL (75)	TPL (78)	TPL (80)	LN (95)	LN (100)	LN (100)	LN (100)
Counterparty-exposure	TPL (82)	TPL (80)	TPL (83)	TPL (80)	LN (94)	LN (99)	LN (100)	LN (100)
Undirected Degree	TPL (94)	TPL (100)	TPL (90)	TPL (100)	SExp (66)	SExp (54)	SExp (52)	SExp (80)
Directed In-Degree	TPL (93)	TPL (94)	TPL (90)	TPL (100)	SExp (58)	SExp (66)	SExp (65)	SExp (60)
Directed Out-Degree	TPL (83)	TPL (89)	TPL (95)	TPL (100)	SExp (64)	SExp (83)	SExp (87)	SExp (100)
Multidirected In-Degree	TPL (92)	TPL (97)	TPL (95)	TPL (100)	SExp (78)	SExp (80)	SExp (87)	SExp (100)
Multidirected Out-Degree	TPL (79)	TPL (93)	TPL (90)	TPL (100)	SExp (67)	SExp (73)	SExp (87)	SExp (100)
In-Exposure	TPL (94)	TPL (97)	TPL (100)	TPL (100)	LN (74)	LN (83)	LN (74)	LN (60)
Out-Exposure	TPL (85)	TPL (94)	TPL (100)	TPL (100)	LN (83)	LN (80)	LN (57)	LN (80)

data coming from the e-mid trading platform. We thus tend to conclude that the evidence on power laws is not (yet) strong enough to warrant their widespread use in policy simulations. From the study presented in this work we provide alternate distributions and corresponding parameters which can systematically and robustly capture the interbank network measures. We deem that those distributions represent a more realistic account of the interbank network structure than the widely used power laws. They could facilitate more realistic contagion modeling and provide more realistic estimates of interbank network measures.

References

[1] Mark EJ Newman. *Power laws, Pareto distributions and Zipf's law*. Contemp. Phys., 46(5):323–351, 2005.

[2] Victor M Yakovenko and J Barkley Rosser Jr. *Colloquium: Statistical mechanics of money, wealth, and income*. Rev. Mod. Phys., 81(4):1703, 2009.

[3] Xavier Gabaix. *Power laws in economics and finance*. Working paper 14299, NBER, 2008.

[4] Rama Cont, Amal Moussa, Edson B Santos, et al. *Network structure and systemic risk in banking systems*, 2013.

[5] Michael Boss, Helmut Elsinger, Martin Summer, and Stefan Thurner 4. *Network topology of the interbank market*. Quant. Finance, 4(6):677–684, 2004.

[6] Daniel O Cajueiro and Benjamin M Tabak. *The role of banks in the Brazilian Interbank Market: Does bank type matter?* Physica A, 387(27):6825–6836, 2008.

[7] Masayasu Kanno. *Assessing Systemic Risk Based on Interbank Exposures in the Japanese Banking System*. Available at SSRN 2482075, 2014.

[8] John Goddard, Hong Liu, Donal Mckillop, and John OS Wilson. *The Size Distribution of US Banks and Credit Unions*. IJEB, 21(1):139–156, 2014.

[9] Giulia De Masi, Giulia Iori, and Guido Caldarelli. *Fitness model for the Italian interbank money market*. Phys. Rev. E, 74(6):066112, 2006.

[10] Giulia Iori, Roberto Reno, Giulia De Masi, and Guido Caldarelli. *Trading strategies in the Italian interbank market*. Physica A, 376:467–479, 2007.

[11] Aaron Clauset, Cosma Rohilla Shalizi, and Mark EJ Newman. *Power-law distributions in empirical data*. SIAM Rev., 51(4):661–703, 2009.

[12] Michael PH Stumpf and Mason A Porter. *Critical truths about power laws*. Science, 335(6069):665–666, 2012.

[13] A.G. Haldane. *Rethinking the Financial Network*, Apr 2009. Speech delivered at the Financial Student Association, Amsterdam.

[14] M. Buchanan. *Meltdown Modelling*. Nature (London), 460(7256):680–682, Aug 2009.

[15] Andreas Krause and Simone Giansante. *Interbank lending and the spread of bank failures: A network model of systemic risk*. J. Econ. Behav. Organ., 83(3):583–608, 2012.

[16] Tarik Roukny, Hugues Bersini, Hugues Pirotte, Guido Caldarelli, and Stefano Battiston. *Default cascades in complex networks: Topology and systemic risk*. Sci. Rep., 3, 2013.

[17] Grzegorz Hałaj and Christoffer Kok. *Assessing interbank contagion using simulated networks*. Comput. Manag. Sci., 10(2-3):157–186, 2013.

[18] Co-Pierre Georg. *The effect of the interbank network structure on contagion and common shocks*. J. Banking Finance, 37(7):2216–2228, 2013.

[19] Alexei Karas and Koen Schoors. *Heracles or Sisyphus? Finding, cleaning and reconstructing a database of Russian banks*. Working paper 327, Ugent, 2005.

[20] Serafin Martinez-Jaramillo, Biliana Alexandrova-Kabadjova, Bernardo Bravo-Benitez, and Juan Pablo Solórzano-Margain. *An empirical study of the Mexican banking systems network and its implications for systemic risk*. J. Econ. Dyn. Control, 40:242–265, 2014.

[21] Giulia Iori, Giulia De Masi, Ovidiu Vasile Precup, Giampaolo Gabbi, and Guido Caldarelli. *A network analysis of the Italian overnight money market*. J. Econ. Dyn. Control, 32(1):259–278, 2008.

[22] Daniel Fricke and Thomas Lux. *On the distribution of links in the interbank network: Evidence from the e-mid overnight money market*. Working paper 1819, IfW, 2013.

[23] Tarik Roukny, Co-Pierre Georg, and Stefano Battiston. *A network analysis of the evolution of the German interbank market*. Technical report, Discussion Paper, Deutsche Bundesbank, 2014.

[24] Ben Craig and Goetz Von Peter. *Interbank tiering and money center banks*. J. Financ. Intermed., 23(3):322–347, 2014.

[25] F Kyriakopoulos, S Thurner, C Puhr, and Stefan W Schmitz. *Network and eigenvalue analysis of financial transaction networks*. The European Physical Journal B-Condensed Matter and Complex Systems, 71(4):523–531, 2009.

[26] Kimmo Soramäki, Morten L Bech, Jeffrey Arnold, Robert J Glass, and Walter E Beyeler. *The topology of interbank payment flows*. Physica A, 379(1):317–333, 2007.

[27] Samantha Cook and Kimmo Soramaki. *The Global Network of Payment Flows*. SWIFT Institute Working Paper, 2014.

[28] Kei Imakubo and Yutaka Soejima. *The transaction network in Japans interbank money markets*. Bank Japan Monet. Econ. Stud., 28:107–150, 2010.

[29] Camelia Minoiu and Javier A Reyes. *A network analysis of global banking: 1978–2010*. J. Fin. Stability, 9(2):168–184, 2013.

[30] Xuqing Huang, Irena Vodenska, Shlomo Havlin, and H Eugene Stanley. *Cascading failures in bi-partite graphs: model for systemic risk propagation*. Sci. Rep., 3, 2013.

[31] Prasanna Gai, Andrew Haldane, and Sujit Kapadia. *Complexity, concentration and contagion*. J. Monetary Econ., 58(5):453–470, 2011.

[32] Erlend Nier, Jing Yang, Tanju Yorulmazer, and Amadeo Alentorn. *Network models and financial stability*. J. Econ. Dyn. Control, 31(6):2033–2060, 2007.

[33] Stefano Battiston, Domenico Delli Gatti, Mauro Gallegati, Bruce Greenwald, and Joseph E Stiglitz. *Default cascades: When does risk diversification increase stability?* J. Fin. Stability, 8(3):138–149, 2012.

[34] Michael Gofman. *Efficiency and stability of a financial architecture with too interconnected to fail institutions*. Available at SSRN, 2012.

[35] Tiago P Peixoto. *Hierarchical block structures and high-resolution model selection in large networks*. Phys. Rev. X, 4(1):011047, 2014.

[36] Benjamin M Tabak, Marcelo Takami, Jadson Rocha, Daniel O Cajueiro, and Sergio RS Souza. *Directed clustering coefficient as a measure of systemic risk in complex banking networks*. Physica A, 394:211–216, 2014.

[37] Duncan J Watts and Steven H Strogatz. *Collective dynamics of smallworldnetworks*. Nature (London), 393(6684):440–442, 1998.

[38] Jari Saramäki, Mikko Kivelä, Jukka-Pekka Onnela, Kimmo Kaski, and Janos Kertesz. *Generalizations of the clustering coefficient to weighted complex networks*. Phys. Rev. E, 75(2):027105, 2007.

[39] M.E.J. Newman. *The structure and function of complex networks*. SIAM Rev., 45(2):167–256, 2003.

[40] Jeff Alstott, Ed Bullmore, and Dietmar Plenz. *powerlaw: a Python package for analysis of heavy-tailed distributions*. PloS one, 9(1):e85777, 2014.

[41] Jean Laherrere and Didier Sornette. *Stretched exponential distributions in nature and economy:fat tails with characteristic scales*. EPJ B, 2(4):525–539, 1998.

[42] Karl Finger, Daniel Fricke, and Thomas Lux. *Network analysis of the e-MID overnight money market: the informational value of different aggregation levels for intrinsic dynamic processes*. Comput. Manag. Sci., 10(2-3):187–211, 2013.

3

Structure and evolution of the world's historical trade patterns

3.1 Abstract

International trade has been an important driver for the development of our modern world, but capturing trade patterns and their change over time continues to prove a daunting task. Painting a detailed picture of historical trade patterns not only puts a high demand on the availability and quality of data, it also begs for an intuitive and succinct way to describe the resulting patterns. To uncover the overall patterns in the data we adopt the complex network perspective and search for signs of globalization, regionalization or a core-periphery structure. After constructing the historical trade integration network spanning the period 1880s to the late 1980s, we use temporal stochastic block models to extract the meso-scale network structure. This SBM methodology makes full use of all available data, takes the time dimension into account and does not make a priori assumptions about the structure of the network. During the first globalization wave (1880-1913), we find a very strong core-periphery structure, where the core is made up of countries like the USA and Great Britain. However, during the interbellum and the second wave of globalization, we see the slow dismantlement of this structure into a hub-and-spoke pattern where regional clusters are linked to a central hub, a sign of increasing regionalization.

3.2 Introduction

As a mathematical framework to study interactions, network theory is a natural candidate to explore international trade and uncover worldwide trade patterns. It should come as no surprise that its tools to uncover meso-scale structures have been used extensively to address questions on globalization [1], regionalization [2, 3] and the core-periphery structure of international trade [4, 5].

The way in which trade networks have been modeled has evolved strongly over time, even within the family of networks where countries are represented by nodes and the edges depict the flow of trade.[1] Initial studies started with the most simple definition of the edges where binary variables indicated the presence or absence of any trade (import or exports) between two countries e.g. [3, 11]. There are many ways in which the edge definition can be more informative on the actual trade flows between the countries. Edges can be directed to distinguish a situation where one country A sells goods to another country B but not vice versa [5, 12–15]. The edges can be weighted with the volume of exports [16–18], allowing us to distinguish countries that trade a lot. These weights can be further refined by making the edges proportional to either the total trade of the country [19] or to the GDP of the country [20], as a trade flow of a million USD will be more important to a small economy that trades little, than to a large country that trades a lot. Finally, the networks constructed in this way can be linked over time, allowing us to study the changing role of certain countries in the trade network, or examine how the overall structure changes over time.

Using these networks social scientists have tried to discern structure of the worldwide trade pattern and track its evolution over time. Most efforts have been focused on the identification of one of three specific network structures: globalization, regionalization and core-periphery. A popular definition of globalization links it to geographical neutrality or the "*death of distance*": a world in which all countries are connected with each other regardless of how far they lie from each other [21]. Regionalization on the other hand is a situation where the trade network is organized into a few regions that trade intensively with each other, but where inter-regional links are sparse. The core-periphery structure, as defined by world-system analysis, speaks of a world where there is a small group of very densely connected countries (the core), surrounded by a large peripheral group of countries that trades mostly with the core and not among themselves.

This chapter outlines how Peixoto's (2015) temporal stochastic block model

[1] Other ways in which trade networks have been modeled focus on transportation infrastructure or the social network among traders or entrepreneurs. Examples of the former include the ORBIS project [6], which reconstructed the Roman Empire's transportation network and linked to the overland trade expansion, and [7] who looked at road networks and overland trade expansion in Sub-Saharan Africa. [8] and [9] analyzed the social network of traders in medieval Genoa and the eighteenth century Indies, respectively, while [10] for example combined both the social and transportation networks.

(SBM) technique can be used to elicit the structure of worldwide trade patterns and track its evolution over time. There are a number of reasons why this technique is particularly effective in addressing this question. To start, it uses all the information contained in the weighted directed network, unlike some of the more coarse-grained alternatives that are limited to unweighed or undirected networks. Secondly, the SBM technique uses the temporal dimension of the data and is able to find the best fitting structure describing the network consistently throughout the entire period studied. Other studies either (arbitrarily) aggregate over several years or only discuss a limited set of years. Finally, while most tests look for a particular structure in the data (e.g. communities or a core-periphery structure), SBM is an agnostic procedure that does not make any a priori assumptions about the structure. The flexibility of this more generic approach allows us to better detect the transitions of the network structure as well as uncover patterns that run counter to expectations.

After discussing how globalization, regionalization and the core-periphery structure are analyzed in complex network theory, section 3.6 sets out to reveal the structure and evolution of the world's trade patterns from the 1880s to the late 1980s. To model the worldwide trade network we use the historical trade integration (*hti*) index developed in [22]. This index captures the level of trade integration between two countries, i.e. the importance of the bilateral import and export flows to each of their economies. Because it combines various historical data sets and uses statistical techniques to address availability problems, the *hti* allows us to construct one of the most replete images of trade relations over this hundred year period (including e.g. colonial trade flows). Using the index, we are able to construct a temporal weighted directed network capturing worldwide trade integration during three consecutive periods: the first wave of globalization (1880-1913), the interbellum (1919-1939), and the second wave of globalization up till the disintegration of the former USSR (1946-1989).

In section 3.3 we first discuss how international trade concepts like globalization and regionalization can be translated to complex network theory, after which stochastic block models are introduced. Finally, sections 3.5 and 3.6 cover the steps in the construction of the historical trade network and discuss our findings.

3.3 Networks and trade patterns

In this first section we discuss three key concepts from the international trade literature and show the link with corresponding concepts in network theory. We discuss globalization, regionalization and the core-periphery idea.

3.3.1 Globalization

Since the 1980s, the term globalization has been used in a myriad of scientific disciplines, each using its own definition(s). Nevertheless, a common denominator to most definitions is the shift of economic transactions from the local toward the global market. In other words, the role of distance as a determinant of trade flows decreases, and countries increasingly trade with partners all over the world. This idea was summarized by [21] as the "*death of distance*". Together with this definition, the economic history literature identifies two important waves of globalization before and after the World Wars and a period of de-globalization in between.

There has been much discussion on the timing of the first phase of globalization. Some authors believe it started at the end of the nineteenth century (e.g. [23, 24]). Others believe that it started from as early as the 1840s (e.g. [25, 26]). Nevertheless, both recognize that the end of the nineteenth century was part of the first globalization wave. With Great Britain in the lead, the mercantilist era was replaced by the idea of a more free trade regime. The European colonizers also imposed this new trading regime on their colonies and even forced independent countries to open up their trade. Technological progress, such as the use of steam engines and the construction of an extensive railway network significantly reduced trading costs. At the same time, the gold standard offered a stable international trading climate [27].

This liberalizing trend was undone by the first World War and the subsequent conference of Versailles which did little to stabilize international relations. The situation was further exacerbated by the Great Depression and the protectionist policies it induced. At the time, the USA took over the leading role in the world economy but failed to further the free trade agenda and could not pull the world economy out of the recession. World War II strengthened the anti-imperialist nationalist and communist states, the disintegrating effect of which lasted till the 1990s. On top of that, globalization was countered by the use of higher tariffs in support of import-substitution policies, mostly by newly decolonized countries [28]. Although the decades following the second World War are entitled as a second wave of globalization, it was mostly a regional evolution limited to Western Europe and Northern America. This was not only a result of disintegrating communist and anti-imperialist states, but also by the protectionist trade politics of newly decolonized countries [28]. Intensification of trade relations took place but were geographically determined and did not extend to the rest of the world. Geopolitical factors were in comparison with the reduction of trade costs during the first globalization wave- the main determinant of the intensifying trade patterns.

In a network context where countries are represented by nodes and their trade by (directed) edges, globalization and the death of distance would be represented by a uniform increase in the number of edges between all nodes. The theoretical

extreme would be a completely connected network where each country trades with all other countries. A number of studies have looked for globalization by studying the network density, but because of data availability issues this research has mostly been focussed on the second half of the 20th century, i.e. the second globalization wave. For example, [1], [5] and [18] all found an increase in the number of edges. Their trade networks was constructed respectively using import data for 1959 and 1975, using import and export data for 1992 and 1998 and using import data for the period 1950 till 2000. All three studies also concluded that the global trade network was far from totally connected. Using the historical trade integration index, [22] were able to compare both globalization waves and found that the importance of distance has significantly dropped since the early 1880s and that once you control for the increase in the number of countries over time the network density increased during both waves.

3.3.2 Regionalization

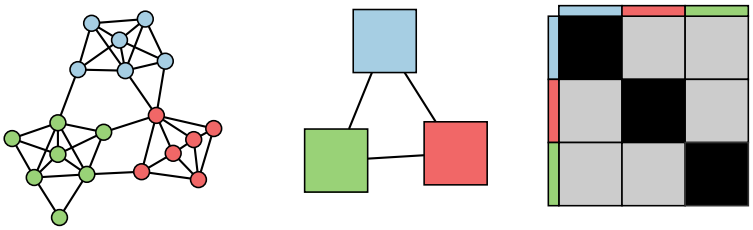
While globalization is an indiscriminate increase in the number of trade links between countries, regionalization implies the emergence of groups of countries that trade strongly amongst themselves, but have few linkages to countries outside of their group. These regions might be actual geographical regions, i.e. determined by distance, but can also be seen more broadly as cultural regions, or those determined by (former) colonial powers and their colonies. Multilateral trade agreements like the European Union or the North American Free Trade Agreements are examples of political institutions aimed at increasing the integration among countries in their region above those in the rest in the world. Whether or not regionalization is a first step to a globalized trade pattern is a hotly debated topic in economics (see e.g. the building block vs stumbling block debate surveyed in [29]).

The concept of trade groups and regionalization has a natural counterpart in network theory: communities and the community-structure. The panels of figure 3.1 show three diagrams representing a typical network with a community structure [30]. First, the left diagram shows the actual network which has three distinct groups of nodes. One can observe that there are many connections among the members of a given group but that there are very little connections to members of other groups. This is the signature of a network with a community structure. The center diagram coarse grains the network and shows a block for each of the corresponding communities. Finally, the right diagram expresses the average number of edges between nodes of a combination of any two groups. For example nodes from block 1 have a high average edge density (black) with other nodes from block 1 but a low edge density (grey) with nodes from block 2 or block 3.[2] In contrast,

[2]Blocks are numbered according to their order of appearance in this edge density matrix. Block 1 is the blue block, block 2 is red, and block 3 is green.

the same figure of a globalized world would be represented by a single block of completely connected countries.

Several studies already investigated the community structure of the international trade network. There are variations in the techniques used as well as the time periods covered. For example [19] study the structure and evolution of the trade network using a visual inspection of so-called "*islands*"[3] starting in 1948. [2] and [3] use a more quantitative technique called modularity maximization [30] to study the trade network respectively starting in 1992 and 1970. While these studies find some evidence of the existence of a community structure,this turns out to be not highly significant. In other words they cannot unambiguously conclude that the worldwide trade network consists of clear-cut, close-knitted communities.



Notes: color online. The network is displayed on the **left**, the **middle** panel shows the coarse-grained view of the network where nodes with a similar position are in the same block, and the **right** panel is a matrix showing the the average edge density between the nodes in each block (darker color = higher density).

Figure 3.1: The community structure-regionalization

3.3.3 Core-periphery

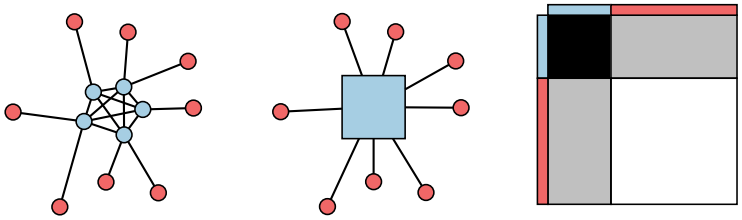
A third structure claimed to be present in international trade is the core-periphery structure. It was first developed by [31]. In his theory the world comprises of a core of industrially developed countries and a periphery of developing countries. The pattern of dependence between the core and periphery is lopsided. The peripheral countries depend predominantly on one of more core countries, while core countries are only connected to other core countries. While Wallerstein's theory also concerns production systems and dependency relations, we will restrict ourselves to the trade pattern associated with the core-periphery model.

If the core-periphery theory holds, we would expect the international trade network to be separable into a core of highly integrated countries and a periphery which is heavily integrated with the core countries but hardly among themselves.

[3] An island is a connected component in which the weights of the arcs within the community are significantly larger than those outside the community. It is a measure implemented in the Pajek library.

The panels of figure 3.2 show the prototypical example of a core-periphery structure using the same three diagrams as above [32]. The tightly connected core countries are represented by the blue nodes while the sparsely connected periphery is formed by the red nodes. The right diagram shows the corresponding edge density fingerprint where the color white signifies an edge density of zero.

[5] looked for this echo in the binary (i.e. unweighted) trade network of the 1990s. Using the Lorenz curves of the degree[4] of the countries in the network, they concluded that the world indeed has a core-periphery structure. [4] used a more complex stochastic block model technique to allocate 118 countries into either the core or the periphery between 1955 and 1970. However, while both studies confirmed the existence of the core-periphery model, they did so using a binary (i.e. unweighted) trade network and covered only a relatively limited number of counties and years.



Notes: the actual network is displayed on the **left**, the **middle** panel shows the coarse-grained view of the network where nodes with a similar position are in the same block, and the **right** panel is a matrix showing the the average edge density between the nodes in each block (darker color = higher density).

Figure 3.2: The core-periphery structure (color online)

3.4 A general descriptor: stochastic block models

In the previous section, we saw that using tests that look for specific network structures, we find evidence that the trade network shows signs of globalization, core-periphery and (to a lesser extent) of regionalization. To some extent, this is because the tests used do not discern well enough between the different possible network structures, either because the test itself is rather crude (e.g. looking at the network density) or because it requires a less informative network (e.g. a binary network). To resolve this problem, we need a method that uses all information available to see which pattern, if any, the worldwide trade network conforms to and how it has changed over time. The solution to this problem was also mentioned in the previous section, namely the so-called stochastic block models (SBM). An SBM is a

[4]The degree of a node in a network is the number of other nodes the node is connected to.

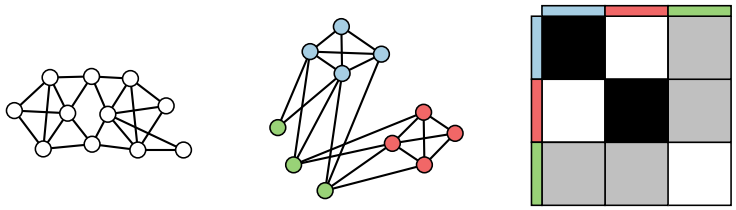
flexible way to provide a high level description of the structure of a network [33]. It identifies groups of nodes that have a similar behavior in the graph, taking into account both incoming and outgoing (weighted) edges. To that end, it summarizes the average relation between the nodes of different partitions. Examples are displayed in the final panels of figures 3.2-3.3. The rows of these matrices show for each block how the outgoing edges are distributed over the different blocks, while the columns show the same for the incoming edges.

Stochastic block models are explicitly designed in order to elicit and describe any type of network structure. It can capture a core-periphery structure, as was done in [4], but it can also distill associative patterns, such as community structures, and dis-associative patterns in the network. Because of their flexibility, picking the right stochastic block model to describe a network is not an easy task and the number of possible partitions increases exponentially with the number of nodes. To pick the optimal partition, we use the methodology developed by [34, 35]. The algorithm looks for the partition that best describes the network with minimum description length, or in other words that uses the least amount of information. In this way, the number of partitions is only increased when it significantly improves the description of the network. Unlike the other methods described above, Peixoto's (2015) method also allows us to take the time-dimension of our dataset into account, allowing us to link different blocks over time and ensuring that the partitioning of the dataset remains consistent and optimal through time.

Lets consider figure 3.3 where we have an example of a general network structure. There are three types of node positions in the network. The blue nodes of block 1 are connected among themselves, are not connected to the red nodes of block 2 and have some connections with the green nodes of block 3.

The nodes of block 2 are also connected among themselves and have some connections with block 3, but are not connected to block 1. Finally, the nodes of block 3, as already said, have some connections with the nodes of block 2 and 3, but are not connected among themselves. Again, we coarse-grain the nodes with a similar structure into blocks in the middle diagram. The right diagram summarizes the average edge density among the nodes of any given pair of blocks.

In summary, this methodology has three particular advantages compared to other techniques. To start, it uses all the information contained in the weighted directed network, unlike some of the more coarse-grained alternatives that are limited to unweighted or undirected networks. Moreover, the SBM technique also uses the temporal dimension of the data and is able to find the best fitting structure describing the network consistently throughout the entire period studied. Other studies either (arbitrarily) aggregate over several years or only discuss a limited set of years. Finally, while most tests look for a particular structure in the data (e.g. communities or a core-periphery structure), SBM is an agnostic procedure



Notes: The actual network is displayed on the **left**, the **middle** panel shows the coarse-grained view of the network where nodes with a similar position are in the same block, and the **right** panel is a matrix showing the the average edge density between the nodes in each block (darker color = higher density).

Figure 3.3: A block-model structure (color online)

that does not make any a priori assumptions about the structure. The flexibility of this more generic approach allows us to better detect the transitions of the network structure as well as uncover patterns that run counter to expectations.

3.5 Constructing the historical trade integration network

In order to uncover the patterns underlying worldwide trade flows, we construct a worldwide network of trade integration using the historical trade integration (*hti*) index from [22]. A detailed discussion of the construction of the *hti* can be found in Appendix A. Trade integration, while based on the bilateral trade flows also takes into account the size of the countries. As a result, it captures the degree of interpenetration, i.e., the extent to which the bilateral import and export flows matter for each country. After all, the importance of a million dollars in trade will be vastly different for Estonia than it will be for the United States, both because of the size of their economy and the total size of their imports/exports. In addition, the dependence on a specific trading partner can be both because you import a lot from or because you export a lot to that country.

The historical trade integration index has a number of advantages over other indicators of trade integration. The most important of which is that it covers a significantly larger fraction of trade, especially in the period before the 1950s. Not only does the index go back to the 1880s (excluding the world wars), it also covers trade between colonies and their colonial power and even trade flows between colonies. While this dataset is far from complete (e.g. it covers only covers the official trade flows), its increased data availability reduces the risk that for example changes in the number of countries covered are mistaken for actual changes in the structure of worldwide trade flows.

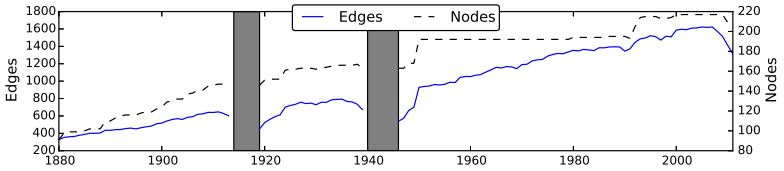
Unlike most other indicators of trade integration the *hti* is a composite index, combining various indicators of trade integration into one overall index. On the one hand, this allows the index to present a more discriminating picture of trade integration. Export and import flows are considered separately, allowing us to identify those countries that strongly depend on one particular market for their imports but not their exports, or vice versa. These trade flows are also normalized in different ways, filtering out important trade flows both because they are large relative to the GDP of the country, or because they constitute a significant fraction of total trade. Because these indicators are combined using a state-space model, differences in their availability can be used in an offsetting manner. Instead of reducing the dataset to those observations where all data is available, the state-space model can compute the index even when only partial data is available. The information on the indicators that are available is combined with past and future values of the index to compute the most likely index value for each observation.[5]

The *hti* does not simply give us one value for the level of integration for each country-pair (dyad) and year, it gives the entire distribution of this value. If the index was computed with only partial data, or if the indicators disagreed on the level of integration, the index of those observations will have a much higher standard deviation. The fact that the entire probability distribution of the index is available is also important for the construction of the trade integration network. First of all, it allows us to construct a much more sparse network by only modeling those edges where the level of integration is significantly greater than zero. Trade flows are marked by an overabundance of very small, but non-zero trade flows.[6] This means that normally either an arbitrary cut-off point has to be chosen, or a much more dense network has to be constructed in which the majority of the edges has a trivial weight. Instead, we construct a weighted directed network by drawing an edge from country A to country B if the $hti_{BA,t}$ is significantly greater than zero at the 1% level. The edge weights are simply the value of the *hti* index.

Figure 3.4 shows the number of nodes and edges of this network over time. The strong increase in the former is for the most part due to the birth of new nations over time, like the Austro-Hungarian Empire disbanding into Austria and Hungary in 1918, or the dissolution of the Soviet Union. Keeping the number of countries constant, we actually see an increase in the network density during 1880-1918 and 1950-2008.

[5]The index can be downloaded at: `http://www.sherppa.ugent.be/hti/hti.html`. Technical details on the construction of the *hti* and a comparison with other indicators of trade integration can be found in [22].

[6]This overabundance of meaningless edges risks significantly slowing down the computation of the more complex network characteristics and analyses.



Notes: number of edges (blue line, right axis) and nodes (dashed line, left axis) in the *hti* network from 1880 to 2015.

Figure 3.4: Number of nodes and edges of the hti network

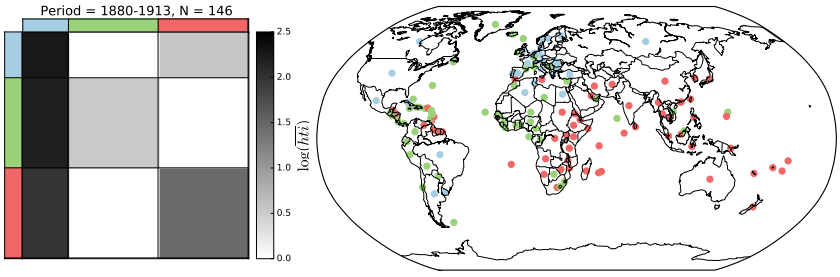
3.6 Revealing the structure of the world's trade patterns

Having constructed the *hti* network, we are now set to elicit the structure of worldwide trade from the 1880s to 1990s. While the SBM technique can be used to analyze how this structure changes from year to year, we will restrict the analysis in this chapter to a more high-level analysis. To that end, we split up the sample into three periods, separated by the world wars: i.e. the first wave of globalization (1880-1913), the interbellum (1919-1939), and the second wave of globalization up till the disintegration of the former USSR (1946-1989).

3.6.1 The first wave of globalization: 1880-1913

We start our discussion with the first wave of globalization, in particular the years 1880-1913. Figure 3.5 shows the results of applying the temporal SBM technique to the *hti* network. The left panel is similar to edge density fingerprint discussed in figure 3.3. The gray scale of square represents the average extent to which the countries in each stochastic block (the rows) is integrated into every other block (the columns). This panel shows that the 146 countries that existed over this period can be separated into three groups. What becomes immediately clear is that the graph is far from symmetric. The first column shows that almost all countries are strongly integrated into the first block (including the first block itself). The other links are far less strong and blocks one and two are poorly integrated with themselves. To aid the interpretation of the fingerprint, the second panel of figure 3.5 shows which countries belong to each of the blocks. As mentioned earlier, only countries which existed and were actively integrated with other countries anytime during the first wave of globalization, will appear in the analysis.[7]

[7]Because borders were subject to change during the years under consideration, we do not use a chloropleth map but use a dot to represent a country. The borders are only there for indicative purposes.



Notes: color online. The **left** panel shows a matrix with the (log of the) average index values of links between the nodes in one block with the nodes of any other block and the **right** panel a world map where countries are colored according to their block assignment.

Figure 3.5: The first globalization wave

Looking at the countries that make up the different blocks in the second panel, we see that the first block (blue) is centered around Western-Europe and North-America although it also includes e.g. Brazil. The second block (green) roughly consists of former Spanish colonies in South-America and the Western part of the African continent, while the third block (red) covers Asia, Oceania and the Eastern part of Africa, lining up more closely with the British colonies at that time.[8] Table 3.1 shows the top five countries according to how much other countries are integrated into them. It shows Great Britain as the top country in the first block -the block which itself has the highest indegree- making it the pivot country during the first globalization wave.

Block 1	Block 2	Block 3
Great Britain	Chile	India
USA	Peru	China
France	Bolivia	Japan
Germany	Ecuador	Austria
Belgium	Columbia	Mauritius

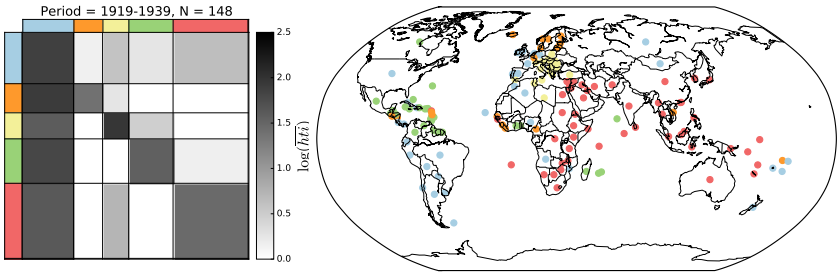
Notes: the top 5 countries per block for the period 1880-1913 according to how strongly other countries are integrated into them.

Table 3.1: Top 5 highest indegree: 1880-1913

[8]The block assignment of some small countries might sometimes seem debatable. Mostly there really is a connection with the block, e.g. France and its overseas territories, but sometimes this is a result of the stochastic nature of the SBM technique. For example we see that for the interbellum Cameroon is assigned to the Scandinavian block. Overall, the bulk of the blocks stays constant when we repeat the partitioning. But sometimes a smaller countries gets reassigned to another block. This reassignment can happen because a country might show behavior which equally likely places it in two different blocks.

In terms of the particular network structure, it is clear that the trade pattern formed during the first globalization wave fits a core-periphery model, where the core consists of the first block (blue) and the periphery can be split up roughly speaking into the former French and Spanish colonies (block 2, green) and the former British colonies (block three, red). All three groups of countries are highly integrated into the core, and the periphery's links to itself are much weaker than their link to the core. This is particularly strong in the case of the French/Spanish colonies, where the link to the core countries is many times stronger than the links to their own peripheral block. However, there is a significant difference between the two groups of peripheral countries. To start, the former British colonies have many more intra-block links than the French/Spanish colonies. Moreover, unlike the latter they also have incoming links from the core countries. Returning to Wallerstein's world-system analysis, we see that while the first block forms the core and the second block the periphery, the third block corresponds to the semi-periphery. Of course, we realize that these findings are not new to the field of economic history. The power of the SBM technique is that it reaches these conclusion using only the raw trade data, without any preconception of the world functioned during this period.

3.6.2 The Interbellum: 1919-1939



Notes: color online. The **left** panel shows a matrix with the (log of the) average index values of links between the nodes in one block with the nodes of any other block and the **right** panel a world map where countries are colored according to their block assignment.

Figure 3.6: Interbellum

The second time period we study is that of the inter-war period. The interbellum is particularly interesting not only because it is known as a period of deglobalization, but also to study the effects of the First World War on the worldwide trade network. To that end, figure 3.6 shows the SBM decomposition of the trade network during the interbellum. The group of countries more or less the same as that during the first globalization wave, although for example end of the Austria-

Hungarian empire resulted in a slight increase in the number of countries (148).

Comparing the pre-WWI to the post-WWI graphs, we see first of all that the number of blocks has increased from three to five. The former group of core countries has split up into three new blocks. This has resulted in a new group of core countries (block 1, blue) that now also includes the newly-independent South-American countries, a second block (orange) consisting mostly of Scandinavian countries and a third group (yellow) of countries from Central and South-Eastern Europe. The fourth group (green) now consists of Canada, Mexico and the Caribbean countries, while the final block (red) remains dominated by the (former) British colonies (table 3.2).

While the pre-WWI trade pattern was an almost perfect example of the core-periphery system, we see that this system has started to shift during the interbellum. While the core group of countries (block 1, blue) still remains the most linked-to block and keeps its pivotal role in the world trade system, most blocks have started forming much stronger intra-block links. The Eastern-European block for example is much stronger connected to itself than to any other block. In addition, the core countries are also forming stronger ties to countries outside of their block. Similarly, the other blocks have started to form links to blocks outside of the core, although these ties remain weak.

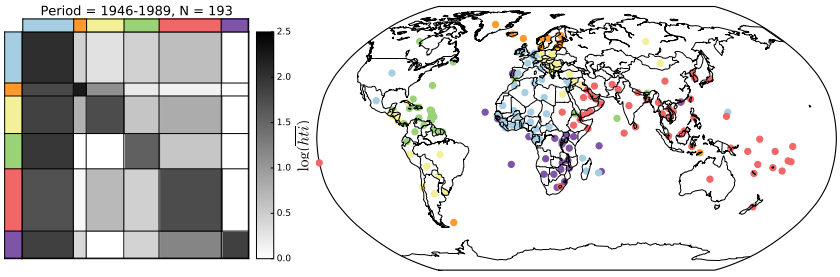
block 1	block 2	block 3	block 4	block 5
Great Britain	Sweden	Italy	Canada	Japan
USA	Denmark	Romania	Virgin Isl. (Br)	India
Germany	Finland	Czechoslovakia	Guyana	Australia
France	Santa Lucia	Austria	Venezuela	China
Belgium	Norway	Greece	Jamaica	Indonesia

Notes: the top 5 countries per block for the period 1919-1939 according to how strongly other countries are integrated into them.

Table 3.2: Top 5 highest indegree: 1919-1939

3.6.3 The second wave of globalization: 1946-1989

Finally, we discuss the second globalization wave from the late 1940s to the the fall of the Berlin wall in 1989. The change of the administrative borders of countries and the birth of new nations that followed the end of the USSR would distort the analysis too much. In comparison to the previous two periods, the number of countries increases with almost a third to a total of 193 in this last period. This is mostly the result of the independence of many North and West-African countries. Even after independence most have remained strongly linked to their former colonial power, which is why the majority of these "*new*" countries are grouped with the former core.



Notes: color online. The **left** panel shows a matrix with the (log of the) average index values of links between the nodes in one block with the nodes of any other block and the **right** panel a world map where countries are colored according to their block assignment.

Figure 3.7: The second globalization wave

The number of blocks needed to describe the network has increased again to six. The biggest change is the split up of the block that was dominated by British colonies (the red blocks in figures 3.5 and 3.6). In general, the blocks now follow mostly regional lines (e.g. the Scandinavian block (block 2, orange) or the Southern Africa block (block 6, purple)), although there are exceptions. For example, as we would expect during the cold-war period, the USSR and the USSR-dominated countries have formed a block separated from the core-group (block 3, yellow). Added to this group is a large contingent of South-American countries, that were grouped together with the USSR countries because of their similar links to a number of Eastern-European countries rather than their link to the USSR. Finally, table 3.3 also shows the United States taking over the leading role in the trade network.

In line with the changes made during the Interbellum, the post WWII period show a further dismantlement of the core-periphery structure. While the first block still remains the most important block and all of the other blocks link strongly to it, almost all blocks link more strongly to themselves. In other words, both the grouping of countries and the trade links between the groups now suggest a more strong regionalized trade pattern. That being said, Western Europe and North America remain an important central player on the world stage, forming links to all other blocks. For example, in spite of their short distance to each other, the two African blocks are linked much more strongly to the core group than to each other. On the whole, we see that with a couple of exceptions, the area of the matrix that is colored white keeps shrinking over time, meaning that most blocks do form trade ties with the other blocks. While this could be interpreted as a sign of an increasingly globalized world, the main trading partners remain either countries from your own block or those from the first block in Western Europe or North America. In other words, rather than a core-periphery pattern seen in the two periods before, the second trade globalization wave follows a hub-and-spoke

pattern where strong regional groupings feed into a central hub. The fact that this pattern contains elements from globalization, core-periphery and communities explains why earlier papers found evidence of these structures.

block 1	block 2	block 3	block 4	block 5	block 6
USA	Canada	Sweden	Russia	Japan	Portugal
Great Britain	Venezuela	Denmark	Romania	Australia	South Africa
Germany	Antilles	Norway	Czech Rep.	India	Zimbabwe
France	Trinidad	Finland	Congo DR	S. Arabia	Kenia
Italy	Guyana	Iceland	Argentina	China	Angola

Notes: The top 5 countries per block for the period 1946-1989 according to how strongly other countries are integrated into them.

Table 3.3: Top 5 highest indegree: 1946-1989

3.7 Conclusion

This chapter studied the structure of the worldwide trade network from the 1880s to the late 1980s, looking specifically for a pattern corresponding to globalization, regionalization or a core-periphery structure. To that end, we constructed a weighted, directed network that captures the extent to which countries are integrated into each other. We did this using the historical trade integration index which expresses for any country-pair worldwide the importance of import and export flows matter on their economy from as early as the 1880s.

The total time-period was split up into three sections: the first globalization wave (1880-1913), the Interbellum, a period of de-globalization (1919-1939) and the second globalization wave until the fall of the Berlin wall (1946-1989). The structure of the network during these periods was analyzed using stochastic block models, specifically using the methodology of [34–36]. These allow us to study all the information in the network (weighted, directed and temporal) and elicit the structure in an agnostic way allowing us to see which of these network structure, if any, underlies in the data.

During the first globalization wave, we find a very strong core-periphery structure, where the core is made up of countries like the USA and Great Britain, the semi-periphery by India and China, and the periphery by Chile and Peru. However, over the next two periods, we see the slow dismantlement of this structure. First of all, an increasing number of countries starts forming its own block and the strength of those intra-block links surpasses in many instances the link to the original core group. The core itself is no longer the same homogenous group it was during the first interbellum as many former colonies are grouped together with their colonial power. In other words, we see signs of increasing regionalization, but unlike the

strict community structure the link with the core group remains strong. While most blocks also start forming links with blocks other than the core (a sign of globalization) these links remain relatively weak. In summary, the start of the second globalization wave is not marked by either globalization, regionalization or core-periphery, but rather a hub-and-spoke pattern where regional clusters are linked to a central hub.

While the current analyses are high level studies that summarize the overall network structure over 20 to 43 years, the technique can also be used to draw the network structure each year and link the different blocks through time. This would allow a much more intricate look at how the network structure has evolved from core-periphery to this hub-and-spoke pattern. For example, it can be used to see whether this change is gradual and happens throughout the period studied, or whether the destabilizing effect of the World Wars is the main contributor to these high level structural changes. A more detailed look would also allow us to flag countries that do not conform to expectations, which could lead to new insights into the behavior of countries on this long time span. We hope to address these and other questions in future research.

References

[1] Sangmoon Kim and Eui-Hang Shin. *A longitudinal analysis of globalization and regionalization in international trade: A social network approach*. Social Forces, 81(2):445–468, 2002.

[2] Matteo Barigozzi, Giorgio Fagiolo, and Giuseppe Mangioni. *Identifying the community structure of the international-trade multi-network*. Physica A: statistical mechanics and its applications, 390(11):2051–2066, 2011.

[3] Javier Reyes, Rossitza Wooster, and Stuart Shirrell. *Regional trade agreements and the pattern of trade: A networks approach*. The World Economy, 37(8):1128–1151, 2014.

[4] David Snyder and Edward L Kick. *Structural position in the world system and economic growth, 1955-1970: A multiple-network analysis of transnational interactions*. American journal of Sociology, pages 1096–1126, 1979.

[5] Raja Kali and Javier Reyes. *The architecture of globalization: a network approach to international economic integration*. Journal of International Business Studies, pages 595–620, 2007.

[6] Walter Scheidel. *ORBIS: the Stanford geospatial network model of the Roman world*. 2015.

[7] Piet Buys, Uwe Deichmann, and David Wheeler. *Road network upgrading and overland trade expansion in Sub-Saharan Africa*. Journal of African Economies, page ejq006, 2010.

[8] Quentin Van Doosselaere. *Commercial agreements and social dynamics in medieval Genoa*. Cambridge University Press, 2009.

[9] Elizabeth Lee Saxe. *Fortune's tangled web: trading networks of English entrepreneurs in Eastern India, 1657-1717*. PhD thesis, 1979.

[10] Emily Erikson and Peter Bearman. *Malfeasance and the Foundations for Global Trade: The Structure of English Trade in the East Indies, 1601–18331*. American Journal of Sociology, 112(1):195–230, 2006.

[11] David A Smith and Douglas R White. *Structure and dynamics of the global economy: network analysis of international trade 1965–1980*. Social forces, 70(4):857–893, 1992.

[12] Ma Ángeles Serrano and Marián Boguñá. *Topology of the world trade web*. Physical Review E, 68(1):015101, 2003.

[13] Xiang Li, Yu Ying Jin, and Guanrong Chen. *Complexity and synchronization of the world trade web*. Physica A: Statistical Mechanics and its Applications, 328(1):287–296, 2003.

[14] Diego Garlaschelli and Maria I Loffredo. *Fitness-dependent topological properties of the world trade web*. Physical review letters, 93(18):188701, 2004.

[15] Diego Garlaschelli and Maria I Loffredo. *Structure and evolution of the world trade network*. Physica A: Statistical Mechanics and its Applications, 355(1):138–144, 2005.

[16] Giorgio Fagiolo, Javier Reyes, and Stefano Schiavo. *On the topological properties of the world trade web: A weighted network analysis*. Physica A: Statistical Mechanics and its Applications, 387(15):3868–3873, 2008.

[17] Giorgio Fagiolo, Javier Reyes, and Stefano Schiavo. *World-trade web: Topological properties, dynamics, and evolution*. Physical Review E, 79(3):036115, 2009.

[18] Luca De Benedictis and Lucia Tajoli. *The world trade network*. The World Economy, 34(8):1417–1454, 2011.

[19] Irena Tzekina, Karan Danthi, and Daniel N Rockmore. *Evolution of community structure in the world trade web*. The European Physical Journal B, 63(4):541–545, 2008.

[20] Gautier M Krings, Jean-François Carpantier, and Jean-Charles Delvenne. *Trade integration and trade imbalances in the European Union: a network perspective*. PloS one, 9(1):e83448, 2014.

[21] Frances Caimcross. *The death of distance*. Harvard Business School Press, Cambridge, 1997.

[22] Samuel Standaert, Stijn Ronsse, and Benjamin Vandermarliere. *Historical trade integration: globalization and the distance puzzle in the long twentieth century*. Cliometrica, pages 1–26, 2016.

[23] Antoni Estevadeordal, Brian Frantz, and Alan M Taylor. *The rise and fall of world trade, 1870-1939*. Technical report, National Bureau of Economic Research, 2002.

[24] Dilip K Das. *Economic Dimensions of Globalization*. In The Economic Dimensions of Globalization, pages 67–102. Springer, 2004.

[25] Kevin H O'Rourke and Jeffrey G Williamson. *Once more: When did globalisation begin?* European Review of Economic History, 8(01):109–117, 2004.

[26] David S Jacks, Christopher M Meissner, and Dennis Novy. *Trade costs in the first wave of globalization*. Explorations in Economic History, 47(2):127–141, 2010.

[27] Nicholas Crafts. *Globalisation and economic growth: a historical perspective*. The World Economy, 27(1):45–58, 2004.

[28] Ronald Findlay and Kevin H O'Rourke. *Power and plenty: trade, war, and the world economy in the second millennium*, volume 51. Cambridge Univ Press, 2007.

[29] Arvind Panagariya. *The regionalism debate: an overview*. The World Economy, 22(4):455–476, 1999.

[30] Santo Fortunato. *Community detection in graphs*. Physics reports, 486(3):75–174, 2010.

[31] Immanuel Maurice Wallerstein. *World-systems analysis: An introduction*. Duke University Press, 2004.

[32] Stephen P Borgatti and Martin G Everett. *Models of core/periphery structures*. Social networks, 21(4):375–395, 2000.

[33] Paul W Holland, Kathryn Blackmond Laskey, and Samuel Leinhardt. *Stochastic blockmodels: First steps*. Social networks, 5(2):109–137, 1983.

[34] Tiago P Peixoto. *Parsimonious module inference in large networks*. Physical review letters, 110(14):148701, 2013.

[35] Tiago P Peixoto. *Hierarchical block structures and high-resolution model selection in large networks*. Physical Review X, 4(1):011047, 2014.

[36] Tiago P Peixoto. *Inferring the mesoscale structure of layered, edge-valued, and time-varying networks*. Physical Review E, 92(4):042807, 2015.

4

Discrete hierarchy of sizes and performances in the exchange-traded fund universe

4.1 Abstract

Using detailed statistical analyses of the size distribution of a universe of equity exchange-traded funds (ETFs), we discover a discrete hierarchy of sizes, which imprints a log-periodic structure on the probability distribution of ETF sizes that dominates the details of the asymptotic tail. This allows us to propose a classification of the studied universe of ETFs into seven size layers approximately organized according to a multiplicative ratio of 3.5 in their total market capitalization. Introducing a similarity metric generalising the Herfindhal index, we find that the largest ETFs exhibit a significantly stronger intra-layer and inter-layer similarity compared with the smaller ETFs. Comparing the performance across the seven discerned ETF size layers, we find an inverse size effect, namely large ETFs perform significantly better than the small ones both in 2014 and 2015.

4.2 Introduction

An exchange-traded fund (ETF) can be thought of as a portfolio of stocks, commodities, or bonds, which is traded like stocks on stock exchanges. Exchange-

traded funds have been made available as investment funds in the US in the early nineties and in Europe in the late nineties. Ever since, ETFs have emerged as a very important investment vehicle attracting ever increasing volumes of capital. Its attractiveness is partly due to the relatively low management and transaction costs involved, an element that is particularly important in times of low yields and low interest rates. Exchange-traded funds represent an increasingly important investment vehicle with potential hazards for systemic risk and possible dangerous menaces for the financial system [1] [2] [3]. For example, it has been shown that arbitrageurs can contribute to cross-sectional return co-movement via ETF arbitrage. The presence of a stock in ETFs increases return co-movement at both the fund and the stock levels, where the effect is strongest among small and illiquid stocks [4]. These days, ETFs come in many different types of flavours [5]. For example, the degree of active management varies very much from one ETF to another.

The focus of this paper is on establishing a taxonomy of the equity ETF landscape on the basis of their size. From our discussion we exclude leveraged ETFs and ETFs holding bonds and commodities, mainly to not overly complicate the analysis. As our focus is on determining the robust and stylized features of the equity ETF landscape using size, we do not segregate by types of ETFs, for example in terms of managed versus active versus passive, or index tracking ETFs.

Size distributions often carry information about the underlying dynamics of a system. The analysis of the distribution of the equity ETF sizes described below discloses some features that suggest departures from a simple power-like tail. The occurrence of a fat tail in the distribution of ETF sizes does not really come as a surprise given the well-documented approximate Zipf-law distribution of firm capitalisations [6]. The fact, however, that there are strong indications that the tail is decorated with some log-periodic structure is remarkable. As this structure is connected with discrete scale invariance, one can infer some interesting constraints on the underlying dynamics of the equity ETF universe. Accordingly, we consider the disclosed log-periodic structure in the size distribution as a natural tool for classification of the universe of ETFs. The inferred classification of the ETFs in several size layers is used to study various economic indicators. We address questions like: 'How similar are the various kinds of ETFs?'; 'How do ETFs distribute their holdings over the wide landscape of possible holdings?'; and 'Is there a connection between the ETF size and their performance?'. These questions are naturally motivated by the existence of the size effect, exploited in the famous Fama-French 3 factor model [7] that also addresses the fundamental issues of the relationships between diversification and performance.

The remainder of this paper is organized as follows. In Section 4.3 we present our empirical analysis of the equity ETF size distribution. We start off (Section 4.3.1) with providing details of the ETF size data used and with performing a

maximum-likelihood fit to their distribution. This reveals indications for an interesting discrete hierarchical structure in the ETF size distribution that is discussed in more detail in Section 4.3.2. In order to put this structure on more solid grounds and to get better hold on the disclosed periodicity in the size distribution, in Section 4.3.3 we pursue a detailed analysis of the ETF size distribution using kernel density estimation and Lomb periodograms. In Section 4.3.4 we sketch some dynamical features of the ETF universe that may give rise to the observed hierarchical structure. We work out in detail how a model based on nonextensive (or, Tsallis) statistical mechanics, a current generalization of Boltzmann-Gibbs (BG) statistical mechanics, can give rise to the discerned oscillatory structures in the ETF size distribution. The basic premises of the proposed model is that the system consisting of all ETFs operates as an open system in a capital reservoir. The size of the ETF system is subject to capital exchange with the reservoir, whereby there is a mechanism of both preferential attachment and growth. In Section 4.4 we introduce a classification into seven layers of the equity ETFs based on the discerned log-periodic hierarchy. We also explore how the economic properties vary over the various size layers. Thereby, we investigate the intra-layer and inter-layer similarities (Section 4.4.1), the variations in the stock holding ubiquity and capitalisation over the different layers (Section 4.4.2), and the connection between layer and performance (Section 4.4.3). Our conclusions are drawn in Section 4.5.

4.3 Analysis of the distribution of ETF sizes

4.3.1 Distribution of total net asset values of ETFs

At the end of 2014, we collected data for all exchange-traded funds (ETFs) labelled as equity ETFs from Thomson Reuters Eikon. This resulted in a set of 479 ETFs for which we obtained the total net assets and the entire composition of their portfolios. In total, this comprised $11,643$ different assets and about $100,000$ positions, for a total net assets over all ETFs of 1.399×10^{12} US$. Figure 4.1 includes the complementary cumulative distribution function (CCDF) of the total net assets of ETFs, i.e., the fraction of ETFs of total net assets larger than or equal to S. Also shown is the CCDF of the log-normal that best fits the data, as obtained by the maximum-likelihood method. The probability density function (PDF) of the log-normal law $\ln\mathcal{N}$ reads

$$\ln\mathcal{N}(\mu_L,\sigma_L^2)=\frac{1}{x\sqrt{2\pi\sigma_L^2}}e^{-\frac{(\ln x-\mu_L)^2}{2\sigma_L^2}}\ , \tag{4.1}$$

with μ_L the location and σ_L the scale parameter whose maximum-likelihood estimates are $\widehat{\mu_L}=18.7$ and $\widehat{\sigma_L}=2.24$. This corresponds to the mode (or most probable) ETF size of approximately 130×10^6 US$ and a mean ETF size of 1.6×10^9

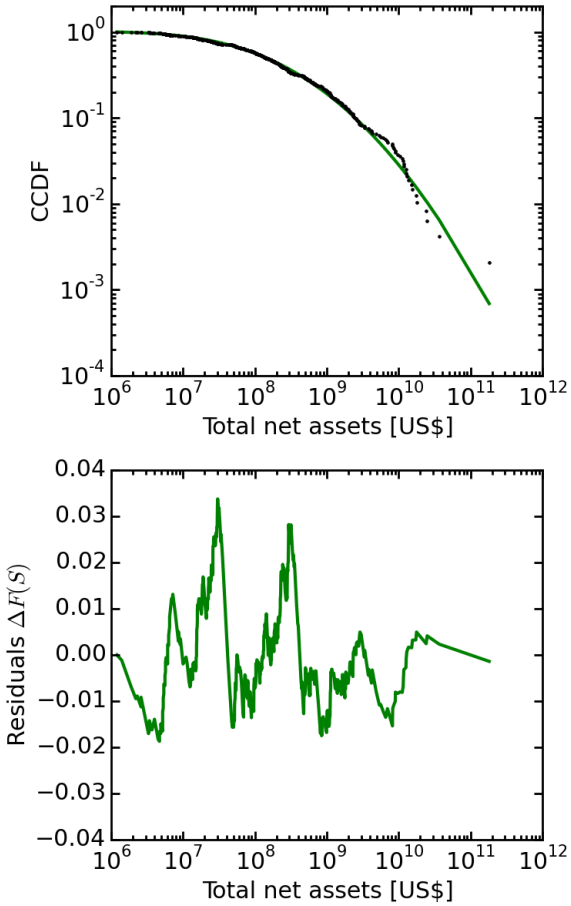


Figure 4.1: The left figure shows the empirical complementary cumulative distribution function (CCDF) of total net assets as a function of total net assets (decimal log-log scale) for the set of 479 equity ETFs collected from Thomson Reuters Eikon in December 2014. The green full line is the maximum likelihood estimation of the lognormal distribution of Eq. 4.1 with $\widehat{\mu_L} = 18.7$ and $\widehat{\sigma_L} = 2.24$. The right figure shows the corresponding residuals – the difference between the lognormal fit and the data – as a function of total net assets.

US$. The much larger value of the mean compared to the mode reflects the existence of a very strong "fat tail" quantified by $\widehat{\sigma_L}$.

When referring to fat tails, it is often convenient to use power law distributions. The tail of a log-normal distribution with large variance (as found here) is difficult to distinguish from a power law distribution (see e.g. Ref. [8] and Section 4.1.3 of Ref. [9]). Indeed, visually, the tail of the empirical CCDF shown in

Fig. 4.1 seems roughly compatible with an asymptotic power law with an exponent of about 1 (Zipf's law). Such an approximate asymptotic Zipf's law has been documented for the distribution of firm sizes [6]. The fact that a similar approximate behaviour in the asymptotic tail is observed for the distribution of ETF sizes is not really a surprise as it can be expected from the presence of two joint and mutually reinforcing mechanisms. First, it is well known that the size of individual firms approximately obeys Zipf's law [6, 10–13]. This result is robust [14] and has been confirmed for different countries [10] and for several measures of firm size including number of employees, profits, sales, value added, and market capitalizations. Therefore, randomly generated portfolios with weights roughly proportional to firm capitalisations will also have an asymptotic Zipf distribution in their tail, as a result of the generalized central limit theorem (see Section 4 of Ref. [9] for a pedagogical presentation). Second, Zipf's law appears quite generically from the combination of three very robust ingredients, namely ETFs are born, they grow via proportional growth and then can also die or close. As outlined in Refs. [15] (Chapter 10) and [16], mergers and acquisitions do not change significantly the overall picture. If the stochastic component of proportional growth is large, Zipf's law is generically an excellent approximation of the tail [15, 17].

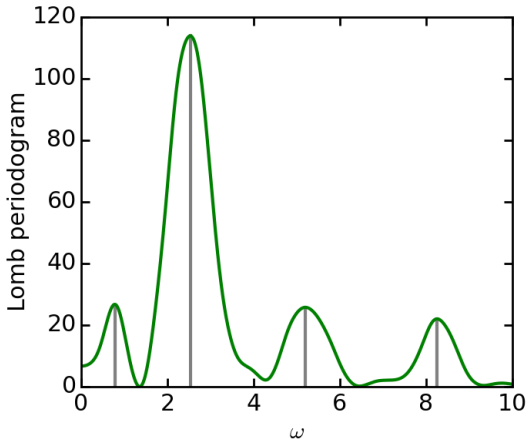


Figure 4.2: Lomb periodogram of the residual function $\Delta F(S)$ shown in the right figure of Fig. 4.1. Here, ω is the conjugate variable to the logarithm of the ETF sizes. The occurrence of the three peaks at $\omega_1 = 2.5$ (large peak), $\omega_2 = 5.2 \approx 2\omega_1$ and $\omega_3 = 8.2 \approx 3\omega_1$ is interpreted in the text. As explained in the text, the peak at $\omega = 0.78$ is likely due to the conjunction of noise in the presence of a finite range of analysis.

Therefore, the observation of a fat tail that looks roughly like Zipf's law is not of much significance. What is much more surprising is the existence of very large deviations from a smooth tail, as made apparent by the structure of the residuals

$\Delta F(S)$ of the lognormal calibration also shown in Fig. 4.1. The pattern of these residuals clearly dominates the question of what is the asymptotic behaviour at large ETF sizes. A first preliminary conclusion is that there appears to be significant more texture to the tail of the CCDF than just a power law or log-normal tail. We now turn to the detailed quantitative analysis of these residuals.

4.3.2 Evidence of a discrete hierarchical texture in the distribution of ETF sizes by spectral analysis of the residuals

A visual inspection of the residuals $\Delta F(S)$ shown in Fig. 4.1 suggests a noisy oscillation. To ascertain the significance of this observation, we calculate the Lomb periodogram of these residuals, shown in Fig. 4.2. The use of the Lomb periodogram, instead of a Fourier transform, is required as a result of the non-even spacing of the pseudo-time variable, namely the logarithm of the total net assets S. Recall that the Lomb periodogram is a method for spectral analysis, which quantifies the contribution of each frequency to a given signal, based on the local least square fit of sine functions to the data [18]. In our case, the signal is the function $\Delta F(S)$ shown in the right panel of Fig. 4.1 expressed as a function of $\ln S$. A statistically significant oscillatory component would mean that $\Delta F(S)$ can be expressed as

$$\Delta F(S) = A + B\cos[\omega \ln S + \phi] + \mathcal{O}\left((\ln S)^2\right) \ , \tag{4.2}$$

where (A, B, ϕ) are three constants and $\mathcal{O}\left((\ln S)^2\right)$ is a second-order residual function of amplitude much smaller than B.

It is important to note that ω is not an angular frequency in the usual sense, as it is the conjugate variable to $\ln S$ and not to S. In other words, as already mentioned, the Lomb spectral analysis is performed in terms of the variable $\ln S$. Thus, the presence of periodicity in the $\ln S$ variable means that the residual function $\Delta F(S)$ is log-periodic in the function S, i.e. it exhibits the symmetry of "discrete scale invariance" [19, 20]. In particular, ω is dimensionless. Fig. 4.2 exhibits an extremely large peak at $\omega = 2.5 \pm 0.2$, which embodies the value of the scaling ratio $p_1 := \exp(2\pi/\omega_1) = 12.3$ for $\omega_1 = 2.5$ and quantifies the ratio of the geometrical series S_n at which the cosine in expression (4.2) is equal to 1 (i.e. $\omega \ln(S_n) + \phi = 2\pi n$, where n is an arbitrary integer). According to extensive simulations in the possible presence of heavy-tailed and correlated noise [21], one can ascertain that this peak at $\omega = 2.5 \pm 0.2$ is statistically highly significant. It expresses the existence of a discrete hierarchy of ETF sizes, roughly spaced according to the ratio $p_1 = 12.3$. Note also the existence of the two smaller peaks at $\omega_2 = 5.2 \pm 0.2 \approx 2\omega_1$ and $\omega_3 = 8.2 \pm 0.4 \approx 3\omega_1$. The presence of these harmonics strengthens the evidence for log-periodicity [22, 23]. The peak at the

lowest value $\omega = 0.78$ corresponds to an oscillation of about the size of the entire range of values, which can be expected just from cumulative noise effect [24] and we thus ignore it.

4.3.3 Generalized derivative and Lomb periodogram of the PDF of ETF sizes

In science, and especially in statistics, it is challenging to prove the absolute reality of an empirical observation. But one can scrutinise the data with a variety of distinct and complementary methods, which altogether may provide confirming evidence of the claimed phenomenon and thus stronger trust in its genuine existence. Because the claim of discrete scale invariance and of a discrete hierarchical structure in the distribution of ETF sizes is rather unexpected and of possible economic importance, we present a detailed analysis of the observed log-periodicity using a completely different methodology, which follows precisely the procedure described in [25, 26]. The procedure has three components: (i) the kernel density estimation (KDE) of the probability density function (PDF) (instead of using the CCDF) of the ETF sizes; (ii) the construction of the generalized (H, q)-derivative of the PDF, and (iii) the calculation of its Lomb periodogram.

Working with the PDF of ETF sizes has the advantage compared with the CCDF of being a local measure of the distribution, hence less prone to the influence of contamination by systematic biases. However, the PDF is more noisy and harder to estimate with limited data. A standard and robust estimation method consists in constructing its kernel density estimator, which is a kind of smoothed histogram. The Gaussian KDE of the PDF of the logarithms $\ln S$ of ETF sizes is defined as

$$\widehat{f}_\sigma\left(\ln S\right) = \frac{1}{N}\sum_{i=1}^{N}\mathcal{N}\left(\ln S - \ln S_i, \sigma^2\right) \ , \tag{4.3}$$

with $\mathcal{N}(0, \sigma^2)$ a zero-mean Gaussian distribution with variance σ^2, and the sum is over the $N = 479$ data points $\ln S_i$. Further, in the context of KDE one refers to σ as the bandwidth. Figure 4.3 shows the KDE of the PDF of ETF sizes for different bandwidths σ. The optimal bandwidth $\sigma_o = 0.22$ is determined with cross validation. Recall that, in cross validation, the model is first fit to part of the data, after which a quantitative metric is computed to determine how well this model fits the remaining data. Obviously, there are strong indications for oscillatory behavior emerging from the KDE analysis of the PDF of ETF sizes. We use the generalized derivative of this function in order to gain a better insight into this oscillatory behaviour.

The generalized (H, q)-derivative of a function $f(x)$ is defined as [27, 28]

$$D_q^H f(x) \equiv \frac{f(x) - f(qx)}{\left[(1-q)x\right]^H} \ , \tag{4.4}$$

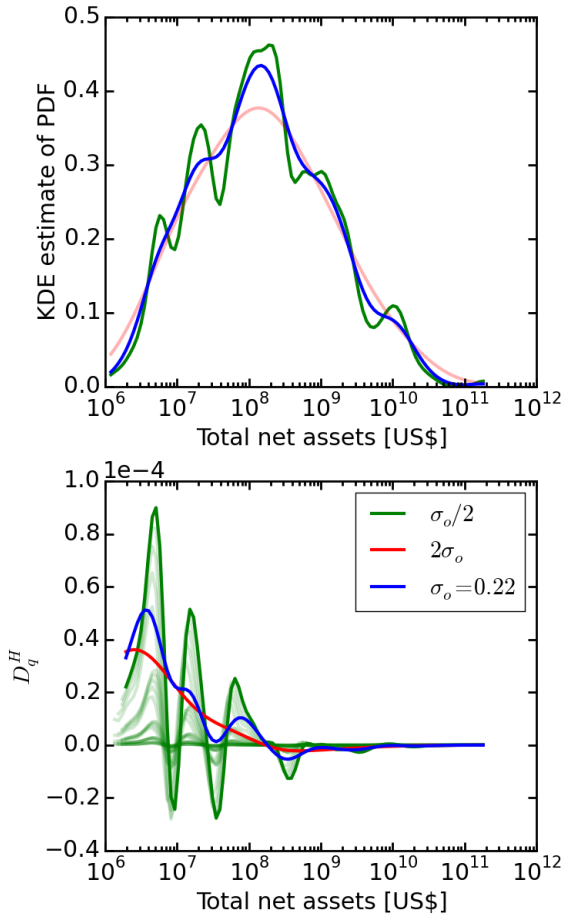


Figure 4.3: The left figure shows the Gaussian KDE of the PDF of the logarithms of the ETF sizes for three different values of the bandwidth. The blue line is for the optimal bandwidth ($\sigma_o = 0.22$) determined using cross validation. The green and red line correspond with a bandwidth of $\sigma_o/2$ and $2\sigma_o$. The right figure shows the generalized derivative $D^{H=0.5}_{q=0.65}$ of the curves of the left figure. For $\sigma_o/2$ we also show D^H_q for different combinations of the values $(0.5 \leq H \leq 0.9, 0.65 \leq q \leq 0.95)$.

and provides a robust metric of the trend or slope of a function. This is particularly useful to detect features in a noisy function, such as the PDF of ETF sizes studied here. Figure 4.3 includes the D^H_q of the KDE of the PDF of ETF sizes for three bandwidths. As recommended in Refs. [27, 28], we have scanned H from 0.5 to 0.9 in steps of 0.08, and q from 0.65 to 0.95 in steps of 0.06 and found that the results are robust. Accordingly, the displayed $D^{H=0.5}_{q=0.65} f(\ln S)$ results can be considered representative. One can observe three to four well formed oscillations

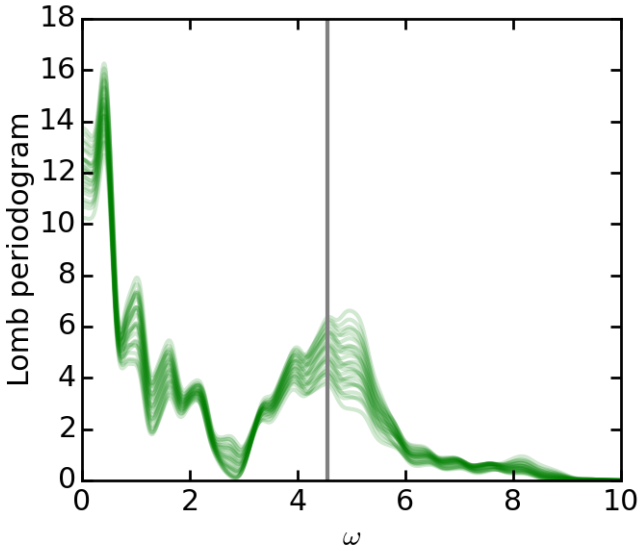


Figure 4.4: Lomb periodogram of the generalized (H, q)-derivative of the Gaussian KDE of the PDF of the decimal logs of the ETF sizes with $\sigma_o/2$ and different values of the combination $(0.5 \leq H \leq 0.9, 0.65 \leq q \leq 0.95)$. The black vertical line is at the center value of $\omega = 4.6 \pm 0.6$, which corresponds to the scaling ratio for the ETF sizes S of $p = \exp(2\pi/\omega) = 3.9 \pm 0.4$.

in the logarithm of the ETF sizes S, quite similarly to the observations of the cumulative distribution approach.

In order to extract the strongest contributing frequencies, we have computed the Lomb periodogram of the generalized (H, q)-derivative of the Gaussian KDE of the PDF of the logarithms of the ETF sizes. We choose the kernel estimation with $\sigma_o/2$ as it is representative of the other estimators but exhibits the largest oscillatory amplitudes. The resulting periodograms are shown in Fig. 4.4. First, the peaks at low angular log-frequencies $\omega < 1.5$ represent oscillations with a wavelength of about the size of the entire range of values and hence can be ignored as explained above [24]. There is only one noticeable peak at a value of $\omega = 4.6 \pm 0.6$ that can be put in correspondence with the second harmonic $\omega_2 = 5.2 \pm 0.2$ previously reported. This angular log-frequency corresponds to a scaling ratio of $p_2 = \exp(2\pi/\omega_2) = 3.9 \pm 0.4$. Note that, when averaging the Lomb periodogram over the scanned H and q, the same estimate $\omega = 4.6 \pm 0.6$ for the unique significant peak is obtained, providing evidence that it has a real existence. There is no significant peak at ω_1, likely as a result of the high-frequency noise associated with the construction of the PDF. Note that the general available theory of log-periodic functions indicates that different harmonics can have very different amplitudes that

depend on subtle properties of the problem [29]. In other words, one should not be surprised that the different harmonics of log-periodicity express themselves with different amplitudes in distinct signals.

4.3.4 Mechanisms of discrete scale invariance in the PDF of ETF sizes

Sections 4.3.2 and 4.3.3 have presented statistically significant evidence of the existence of a discrete hierarchical structure in the distribution of ETF sizes, with preferred scaling ratios approximately equal to $p_1 \approx 12$ and $p_2 = \sqrt{p_1} \approx 3.5$. Reference [19] provides a review of the many mechanisms that can produce such a discrete hierarchy. While we cannot offer a definite mechanism and test for its relevance, the most likely candidates are the Kesten process [30, 31] and aggregation/fragmentation dynamics [32]. The Kesten process can be used to describe the growth of portfolio as a result of the joint addition of new deposits and of stochastic proportional growth. Log-periodic PDFs emerge quite robustly if the multiplicative stochastic factors are not too broadly distributed. The aggregation/fragmentation dynamics could be also a limiting process for the formation of ETF portfolios, for which it can be shown that discrete scale invariance may emerge for quite general aggregation/fragmentation kernels.

Another explanation for the occurrence of distributions with a power law decorated by log-periodic oscillations finds its origin in the formalism of nonextensive statistical mechanics [33]. At any instant of time, the universe of ETFs behaves as an open system that seeks to find equilibrium with the whole of the capital market that acts as a reservoir. In the context of equilibrium statistical physics, this equilibrium process [34] gives rise to the well-known Boltzmann-Gibbs distribution of the sizes of the ETFs

$$P^{BG}\left(0 \leq S \leq \infty\right) = \frac{1}{T_0} \exp - \frac{S}{T_0} \, , \tag{4.5}$$

where the temperature T_0 acts as a typical scale parameter for the size of the ETFs. In this picture, all ETFs (independent of size) are subject to a similar stochastically driven capital exchange with the reservoir of the complete market. In other words, the universe of ETFs is embedded in the reservoir of investment products and the exchange between any element in the ETF universe and the reservoir can be parametrized by a single scale parameter T_0. The larger T_0 the larger the average size of the ETFs. Obviously, Eq. (4.5) does not give rise to fat tails in the distribution of the ETF sizes and cannot be considered realistic given the observations of Section 4.3.1. A generalization, however, proceeds as follows. The Boltzmann-Gibbs exponential distribution (4.5) is a solution to the following equation

$$\frac{dP^{BG}(S)}{dS} = -\frac{1}{T_0} P^{BG}(S) \, . \tag{4.6}$$

In nonextensive statistical mechanics, this equation is extended by adding a nonextensivity parameter n

$$\frac{dP(S)}{dS} = -\frac{1}{T(S)}P(S) = -\frac{1}{T_0 + \frac{S}{n}}P(S) \; . \tag{4.7}$$

A particular solution to this equation is known as the Tsallis distribution

$$P(S) = \frac{n-1}{nT_0}\left(1 + \frac{S}{nT_0}\right)^{-n} \; . \tag{4.8}$$

The Tsallis distribution nicely interpolates between the Boltzmann-Gibbs exponential for ETF sizes S smaller than the scale parameter T_0 and a power law tail for $S \gg T_0$. The nonextensivity parameter n makes the temperature –in the current context a proxy for the typical amount of capital exchange of an ETF with the reservoir of investment products– dependent on the actual ETF size S. Loosely speaking, the parameter $\frac{1}{n}$ can be interpreted as a measure for the degree of preferential attachment [35], or the extent to which the rich ETFs get richer. In the current context, the n accounts for the fact that there is an increased linear tendency of an ETF to accrue money from the reservoir of investment products as it increases in size S. The quantity n determines the asymptotic behaviour of the distribution $P(S)$. The normalization condition $\int_0^\infty P(S)dS$ of the distribution (4.8) requires that $n > 1$. The Tsallis distribution, for example, provides an excellent fit to transverse momentum distributions in high-energy collisions with values of n of the order 6-8 [36]. Obviously, the limit $n \to \infty$ corresponds to a vanishing preferential attachment effect. The smaller n the larger the difference between the temperature associated with the small and the large ETFs. Small ETFs, that are defined as those with a current size smaller than the scale T_0 experience a temperature $T \approx T_0$ in their interaction with the capital market. Large ETFs, defined as ETFs larger than the scale T_0, experience a temperature $T(S)$ that scales linearly with their size $T(S) = T_0 + \frac{S}{n}$.

Building on the connection between preferential attachment growth and nonextensive statistical mechanics [35] and following the derivations of Ref. [20] we now explain that the differential equation of the type (4.7) can give rise to distributions $P(S)$ that have a power law tail decorated with log-periodic oscillations if one adds an evolutionary aspect to the system. In finite difference form, the Eq. (4.7) can be written as

$$P(S + \delta S) = \frac{-n\delta S + nT_0 + S}{nT_0 + S}P(S) = \frac{-\delta S + T(S)}{T(S)}P(S) \; , \tag{4.9}$$

where δS can be interpreted as a single-step small increment of the ETF size S. We now seek to find the solutions to the evolution equation (4.9) for a specific choice for the increment δS. As the changes δS can be anticipated to be proportional to

the fluctuating temperature $T(S)$ one can introduce an additional scale parameter γ

$$\delta S \equiv \gamma n T(S) = \gamma n \left(T_0 + \frac{S}{n} \right) = \gamma n T_0 \left(1 + \frac{S}{nT_0} \right) \ , \tag{4.10}$$

where γ can be made arbitrary small by imposing the condition $\gamma \ll \frac{1}{n}$ and recalling that $n > 1$. After inserting the expression (4.10) into (4.9) one finds that

$$P\left(S(1+\gamma) + \gamma n T_0\right) = (1 - \gamma n) P(S) \ . \tag{4.11}$$

In the asymptotic regime $S \gg T_0$, one finds

$$P\left(S(1+\gamma)\right) \approx (1 - \gamma n) P(S) \quad (S \gg T_0), \tag{4.12}$$

an expression that for finite values of γ is directly recognized as the usual condition $P(\lambda S) = \mu P(S)$ for scale invariance of the function $P(S)$. It is well known [19, 20] that the most general solution for the asymptotic part of the distribution is a linear combination of power laws with complex exponents $\alpha_k(\gamma, n)$

$$P(S) \approx \sum_{k \in \mathbb{N}} w_k S^{-\alpha_k(\gamma,n)} \quad (S \gg T_0), \tag{4.13}$$

with,

$$\alpha_{k \in \mathbb{N}}(\gamma, n) = -\frac{\ln(1 - n\gamma)}{\ln(1+\gamma)} + \frac{2\pi i k}{\ln(1+\gamma)} \ . \tag{4.14}$$

As is usually done, we retain only the terms in w_0 and w_1 and the real part of the function, to obtain

$$P(S) \sim S^{-n - \frac{n}{2}(n+1)\gamma + \mathcal{O}(\gamma^2)} \left[w_0 + w_1 \cos\left(\frac{2\pi}{\ln(1+\gamma)} \ln S \right) \right] \quad (S \gg T_0) \ . \tag{4.15}$$

For large values of the ETF size S ($S \gg T_0$), the distribution $P(S)$ behaves as a power law decorated with a log-periodic oscillation of the type $\Delta F(S)$ defined in Eq. (4.2). This is compatible with the qualitative findings for the tail parts of the empirical distribution of ETF sizes (see Figs. 4.1 and 4.3). We stress that the log-periodic oscillation in the above distribution $P(S)$ is determined by the finite parameter γ that is connected with the time evolution of the system in accordance with multiplicative size increments δS that obey the relation (4.10). For infinitesimally small increments – that correspond with $\gamma \to 0$ – one has that

$$\lim_{\gamma \to 0} e^{-\alpha_k(\gamma,n)} = e^{-\alpha_0} \quad (\forall k) \ , \tag{4.16}$$

and the asymptotic distribution of (4.15) reduces to the tail S^{-n} of the Tsallis distribution (4.8).

As a matter of fact, the proposed asymptotic solution (4.15) of the evolution equation, provides one with a prediction for the angular frequency of the oscillations in $\ln S$ after one time step

$$\omega_1 = \frac{2\pi}{\ln(1+\gamma)} \ . \tag{4.17}$$

The measured distribution of ETF sizes is the result of many multiplicative evolution steps of the type (4.9) each with its finite characteristic scale parameter γ_t. The size of the ETF at the time instances t and $t - \Delta t$ are connected by an expression of the type (4.10)

$$\delta S_t = S_t - S_{t-\Delta t} = \gamma_t n \left(T_0 + \frac{S_{t-\Delta t}}{n} \right) \ . \tag{4.18}$$

For the sake of simplicity, let us assume that there are κ time steps Δt and that all γ_t are equal: $\gamma_t = \gamma, \forall t$.

Proceeding in a fashion analogous to the above derivations and detailed in Ref. [20], one finds after κ time steps an asymptotic distribution $P(S)$ that is similar to the result of (4.15) apart from the following substitution in the angular frequency ω_κ of the $\cos(\omega \ln S)$ term

$$\omega_1 = \frac{2\pi}{\ln(1+\gamma)} \Longrightarrow \omega_\kappa = \frac{2\pi}{\kappa \ln(1+\gamma)} \ . \tag{4.19}$$

This means that the angular frequency of the oscillations in $\ln S$ decrease with the finite number of time steps κ as $\frac{1}{\kappa}$. As a consistency check and referring to the observed oscillations in the size distributions of the ETFs in Fig. 4.4: for $\omega = 4.6$ one finds $\gamma = 0.014$ for $\kappa = 100$ and $\gamma = 0.0014$ for $\kappa = 1000$.

4.4 Analysis of the economic significance of the hierarchy of ETF sizes

We now explore whether the discrete hierarchical structure in the distribution of ETF sizes could be associated with economic properties of the ETF portfolios. Studying the return-risk properties of the hierarchy of ETF sizes is tantamount to investigating the generalisation of the size factor (also often referred to as SMB for "Small [market capitalization] Minus Big [market capitalization]") of the three-factor Fama-French model [37]. Recall that the introduction of the SMB factor was motivated by the observation that small capitalisation stocks have tended to do better than the market as a whole. The observation that the size effect is rather weak, especially in the last decade, has pushed Fama and French to extend their three-factor model to a five-factor model [38]. Therefore, we expect to find only

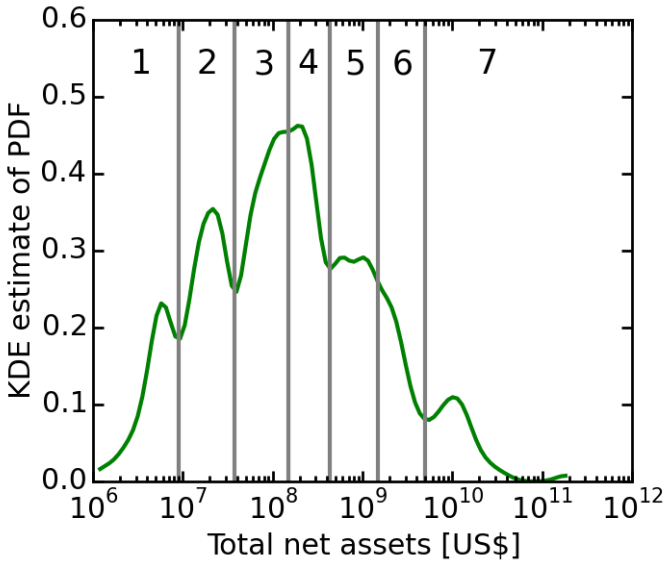


Figure 4.5: Partitioning of the distribution of ETF sizes obtained with the Gaussian KDE of the PDF of the logarithms of the ETF sizes with $\sigma_o/2$, by identifying the minima and maxima that are separated by a factor close to $p_2 \approx 3.5$. The seven size layers are bracketed by the vertical lines.

weak signatures of the size hierarchy. Nevertheless, we propose that it is worthwhile to investigate a generalisation of the dichotomy between small and big ETF sizes, by using the discrete hierarchy discovered above.

In order to construct the size layers partitioning the ETF universe, we use a specific geometric partitioning of the ETF universe based on the discovered scaling ratio $p_2 = \exp(2\pi/\omega_2) \approx \sqrt{p_1} \approx 3.5$, because it is present both in the analysis of the CCDF (Section 4.3.2) and of the PDF (Section 4.3.3) of ETF sizes. Moreover, it amounts to the simplest substructure to the dominant scaling ratio $p_1 \approx 12$ identified in Section 4.3.2. We partition the distribution of ETF sizes obtained with the Gaussian KDE of the PDF of the logarithms of the ETF sizes with $\sigma_o/2$, by identifying the minima and maxima that are separated by a factor close to $p_2 \approx 3.5$. The obtained set of seven size layers are represented in Fig. 4.5. Table 4.1 reports a number of properties for each size layer i, including the number of ETFs, the average number of holdings per ETF, the upper bound size (ub_i) and the ratio ub_i/ub_{i-1}. One can observe that the mean value of this ratio is 3.6, which is close to the scaling ratio of $p_2 = 3.5 \pm 0.2$, as expected. Note that the most probable ETF size of approximately 130 million US$ determined in Section 4.3.1 falls close to the boundary between the third and fourth size layer. In contrast, the mean ETF size of 1.6 billion US$ is close to the boundary between the fifth and sixth size layer.

4.4.1 Intra-layer and inter-layer similarity of stock holdings across ETF size layers

To investigate whether there is a connection between the different size layers and the portfolio composition of the ETFs, we compare the portfolio similarity of the different scales. The portfolio similarity $sim_{ee'}$ of ETFs e and e' is defined as

$$sim_{ee'} = sim_{e'e} = \frac{\sum_{i \in |h_e \cap h_{e'}|} w_{ei} w_{e'i}}{\sqrt{\sum_{k \in h_e} w_{ek}^2} \sqrt{\sum_{l \in h_{e'}} w_{e'l}^2}}, \tag{4.20}$$

where h_e and $h_{e'}$ are all the holdings of ETFs e and e', and w_{ei} is the portfolio weight of holding i in ETF e.

Figure 4.6 represents the matrix of average intra-layer and inter-layer similarities of ETFs, as defined by (4.20) across the seven size layers. Specifically, an entry SIM(i, j) of this matrix is the average similarity $sim_{ee'}$ between the portfolios of all ETFs e in size layer i with all the ETFs e' in size layer j. Firstly, one sees that size layers consisting of larger ETFs are more self-similar. As the size layer number i increases, there is less diversity in the number of holdings used to construct the portfolios of the corresponding ETFs. The SIM$_I \equiv$SIM(I, I) column of Table 4.1 reports these intra-size layer similarities. To sum up, two size layers of large ETFs are more similar than two size layers of smaller ETFs or than a size

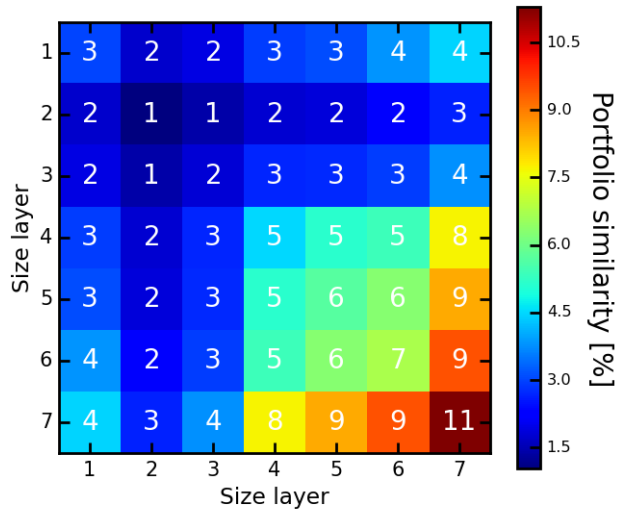


Figure 4.6: Matrix of the average intra-layer and inter-layers similarities of ETFs across the seven size layers defined from Fig. 4.5 and detailed in Table 4.1. An entry SIM(i, j) of this matrix, as indicated by the white number and the color scale, is the average similarity expressed in percentages between the portfolios of all ETFs in size layer i with all the ETFs in size layer j. The similarity between two ETFs is defined by Eq. (4.20).

layer of small ETFs and a size layer of large ETFs. The first smallest size layer 1 breaks this regularity, which is kind of an oddity that can perhaps be associated with the very small sizes of these ETFs. The column SIM_M of Table 4.1 reports the average overlap of the ETFs in a given size layer with the market portfolio. To construct the market portfolio, we consider all the stocks held by ETFs. The weight of a stock in this portfolio is simply its market capitalization divided by the total market capitalization of all the stocks combined. Not surprisingly, one can observe that the larger size layers exhibit a stronger similarity to the market portfolio.

Size layer $]lb_i, ub_i]$ $\left(10^6\text{US\$}\right)$	#ETFs	$\bar{N}_h$	ub_i/ub_{i-1}	SIM_I (%)	SIM_M (%)
1: $]0, 9]$	48	172		3.03	13.04
2: $]9, 38]$	88	150	4.2	1.04	7.78
3: $]38, 150]$	109	183	3.9	1.82	8.54
4: $]150, 430]$	84	229	2.9	4.5	14.37
5: $]430, 1500]$	77	258	3.4	5.72	17.58
6: $]1500, 5000]$	43	281	3.3	6.74	19.44
7: $]5000, \infty]$	30	288		11.29	26.13

Table 4.1: For each identified ETF size layer $i = 1, 2, \ldots, 7$, this table reports the corresponding interval of covered market capitalisations with the upper (ub_i) and lower bound (lb_i). For example, size layer 2 contains ETFs with capitalisations between 9×10^6 US\$ and 38×10^6 US\$. Further, for each size layer we provide the number of ETFs, the average number $\bar{N}_h$ of holdings per ETF, the ratio ub_i/ub_{i-1}, the average overlap similarity over all ETF pairs in a band (SIMI) and the average overlap with the market portfolio (SIM_M).

The greater intra-layer and inter-layer similarity of ETFs of large sizes is not surprising, as a large amount of capital to invest needs to find a large number of potential firms with not too large weights in order to limit market impact. As the universe of available stock investment is finite, and the set of attractive stocks is even more limited at any given time, it can be expected that the large ETFs exhibit significant overlaps in their holdings. For instance, who would not hold the largest firms such at Apple in their portfolio?

To quantify further this similarity in the holdings of the large ETFs, we define the two adjacency matrices $\mathcal{M}_{bh}^{bin}$ and $\mathcal{M}_{bh}^{frac}$ with dimensions $(N_b \times N_h)$. Here, $N_b = 7$ is the number of size layers and $N_h = 11,643$ is the total number of distinct holdings over the 479 equity ETFs considered. The fact that one ETF in the size layer b has a position in stock h is encoded by $\mathcal{M}_{bh}^{bin} = 1$, otherwise $\mathcal{M}_{bh}^{bin} = 0$. The second matrix is defined such that the element $\mathcal{M}_{bh}^{frac}$ is equal to the fraction of ETFs in the size layer b that have a position in the holding h.

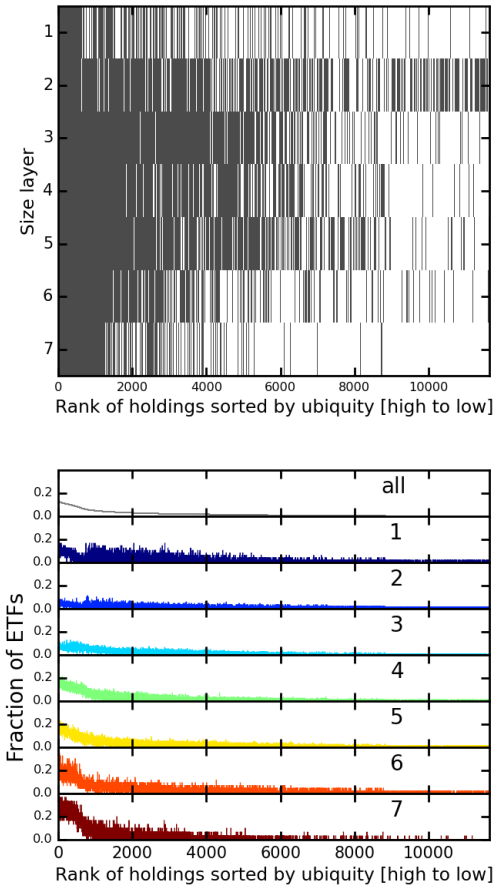


Figure 4.7: The matrices $\mathcal{M}_{bh}^{bin}$ (left) and $\mathcal{M}_{bh}^{frac}$ (right) as defined in the text. The upper panel of $\mathcal{M}_{bh}^{frac}$ shows the fraction of the equity ETFs in which a certain holding occurs. The 11,643 holdings are sorted according to their ubiquity with rank 1 corresponding to the most ubiquitous stock.

The adjacency matrices $\mathcal{M}_{bh}^{bin}$ and $\mathcal{M}_{bh}^{frac}$ are shown in Fig. 4.7. The holdings are sorted from highest to lowest ubiquity in the 479 equity ETFs considered in our analysis. The adjacency matrices of Fig. 4.7 allow us to draw several conclusions. First, larger ETFs tend to use a smaller set of stocks to invest in. Second, larger ETFs tend to select increasingly from the same ubiquitous stocks. The ETFs in size layer 2, on the other hand, nicely sample from the entire space of holdings. Third, the first size layer 1 is an exception to this stylized picture and appears to be a smaller version of size layers 5 and 6.

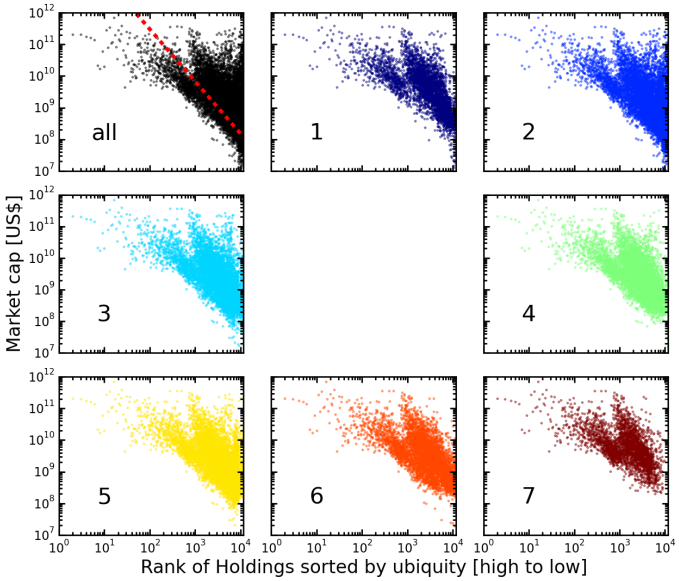


Figure 4.8: For all the 479 ETFs and for each size layer separately, the market capitalization at the end of 2014 for all the holdings considered in Fig. 4.7. The holdings are sorted from highest to lowest ubiquity and the results are plotted on log-log axes (base 10). The dashed red line in the upper left panel indicates the separation between the two clusters mentioned in the text.

4.4.2 Relationship between stock holding ubiquity and capitalisation within ETF size layers

As already mentioned, an obvious explanation for why larger ETFs tend to be more similar is that they need to hold stocks with a larger market capitalisation. Holding too many smaller stocks might prove too costly and not sufficient to absorb the capital in need of investment opportunities. The concentration and similarity of large ETFs may, in large part, just reflect the reduction in available large stocks to invest in. To investigate this hypothesis, we first study the relation between the ubiquity of a stock and its market capitalisation. The results are shown in Fig. 4.8 for all holdings and per size layer respectively. The figure exemplifies that, up to the 500 largest firms (up to rank 500), the sizes of the corresponding firms are drawn from approximately the same distribution with a minimum size of about 5×10^9 US$. In contrast, below rank $\approx$500, one can observe a simple power law relationship relating the size of the smallest admissible firms with respect to their

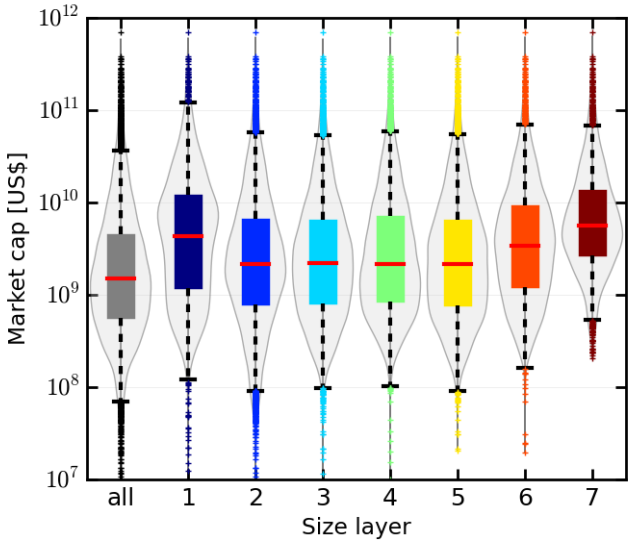


Figure 4.9: For the entire sample of 479 equity ETFs and for each size layer separately, this figure shows the boxplots of the market capitalisations of all the stocks held by the ETFs. We note that the boxplot covering all the holdings is not a mere aggregation of all the size layers. In addition to presenting the boxplots, we show also the so-called violin plots that give the full distribution in thin lines along each vertical axis.

abundance in the universe of ETFs. The corresponding exponent of the power law is $\alpha = 1.7028 \pm 0.0002$ where the power law is $f(x) \propto 1/x^{\alpha+1}$.

As expected, smaller stocks tend to be less ubiquitous in large ETFs. This effect is most obvious in the panel of size layer seven. On the contrary, we do not see that small ETFs only hold small cap stocks. Surprisingly, Fig. 4.8 uncovers some additional structure. There appear to be two clusters in the considered "market cap"-"holding ubiquity" matrix, and to guide the eye we have drawn a separation line in the panel including all stocks.

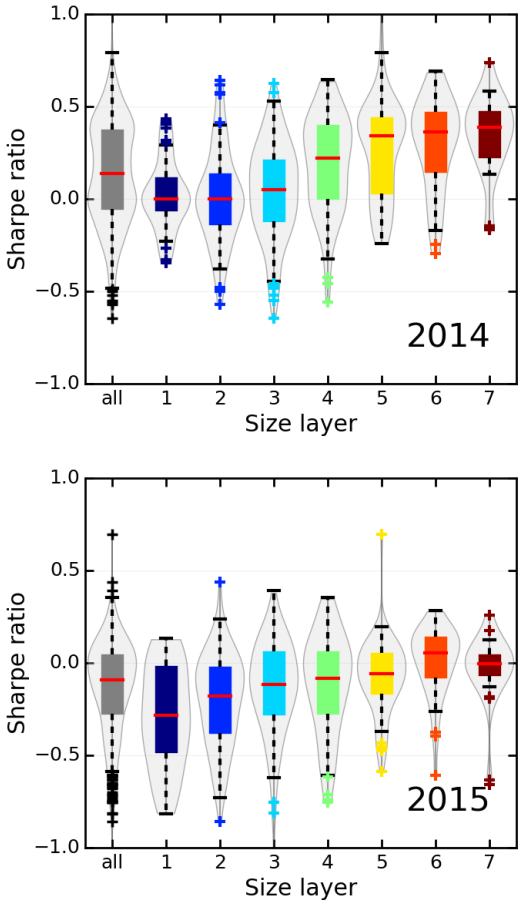


Figure 4.10: Performance for the entire sample of ETFs and the different size layers in the years 2014 (left) and 2015 (right). We present the performance measure as boxplots complemented by violin plots giving the full distribution in thin lines along each vertical axis.

We have also investigated the distribution of the market capitalisation of all

the stocks that appear in the discerned size layers. Figure 4.9 shows the market capitalisations of all the stocks that are held by the ETFs in the considered subset, be it the entire sample of ETFs or a particular size layer. The boxplot for all the holdings over all ETFs is not a mere aggregation of the boxplots of all size layers, as stocks can appear only once. For size layers 6 and 7, but remarkably also for size layer 1, the distribution is clearly shifted to stocks with a larger market capitalisation. The violin plots of the full PDFs for the whole ETF market and each size layer reveal that the distributions have more structure than being uni-modal and symmetric. They exhibit significant skewness as well as bi-modality as is the case for size layers 1 and 5.

4.4.3 Investment performance across the seven ETF size layers

Figure 4.10 presents a measure of performance of the ETFs in each size layer and in total for 2014 and 2015. Unfortunately, we do not have access to comparable data on ETFs for other calendar years, which prevents us from performing factor regressions as in [7, 38]. As metric of performance, we use the Sharpe ratio, defined as the annualised mean return divided by annualised volatility (standard deviation of the returns). We take a reference risk free interest rate equal to 0. In addition to presenting the standard boxplots, we show also the violin plots that give the full distribution in thin lines along each vertical axis.

In 2014, all median Sharpe ratios were positive, showing that the majority of ETFs generated positive returns. In addition, a significant inverse-size effect can be observed: the upper size layers 5-7 significantly over-perform the lowest size layers 1-3, with size layer 4 representing an intermediary case. In contrast, 2015 has been a difficult year for ETFs as well as for hedge-funds in general. Except for size layer 6, all other size layers have negative median Sharpe ratios. One can also observe an inverse size effect, in the sense that the upper size layers also perform better than the lower ones, but the difference is less pronounced than in 2014. The higher performance of large ETFs is reminiscent of the increasing returns to wealth inequality found, for example, for university endowments [39] and for households' portfolios [40]: larger endowments provide much larger returns as a result of better economies of scales, and of access to more investment opportunities and to more skilled managers. Similarly, larger households' wealth enable access to more diversified portfolios.

The violin plots of the full PDFs for the whole ETF market and each size layer in Fig. 4.10 reveal that the distributions have more structure than being uni-modal and symmetric. The discerned bi-modality of the PDF of Sharpe ratios for the whole market can be attributed to the distinct performance of two classes of ETFs, the less performing one represented mostly in size layers 2-4 and the more performing one populating the size layers 5-7.

4.5 Conclusion

We provided a novel detailed analysis of the size distribution of a universe of almost 500 equity ETFs and discovered a discrete hierarchy of sizes, which imprints a log-periodic structure on the probability distribution of ETF sizes that dominates the details of the asymptotic tail. We used the found discrete hierarchy to propose a classification of the whole universe of ETFs into seven size layers. Introducing a similarity metric, we found that the largest ETFs exhibit both stronger intra-layer similarity and stronger inter-layer similarity compared with the smaller ETFs. We have found strong indications that this reflects the obligation for large ETFs to spread their capitalisation on a relatively more reduced set of large stocks. This lack of diversification in the classes of large ETFs seems to reinforce the concentration of stock capitalisation known as Zipf's law. This concentration together with the similarity of holdings suggests potential vulnerability to systemic risks. We also provided comparative performance across the seven ETF size layers and found an inverse size effect, namely large ETFs perform significantly better than the small ones.

References

[1] I Diaz-Rainey and G Ibikunle. *A Taxonomy of the Dark Side of Financial Innovation: The Cases of High Frequency Trading and Exchange Traded Funds*. International Journal of Entrepreneurship and Innovation Management, 16(1/2):51–72, 2012.

[2] Ayan Bhattacharya and Maureen O'Hara. *Can ETFs Increase Market Fragility? Effect of Information Linkages in ETF Markets*. Available at SSRN: http://ssrn.com/abstract=2740699, 2016.

[3] Semyon Malamud. *A Dynamic Equilibrium Model of ETFs*. Swiss Finance Institute Research Paper No. 15-37. Available at SSRN: http://ssrn.com/abstract=2662433 or http://dx.doi.org/10.2139/ssrn.2662433, 2015.

[4] Zhi Da and Sophie Shive. *When the bellwether dances to noise: Evidence from exchange-traded funds*. Available at SSRN 2158361, 2013.

[5] https://en.wikipedia.org/wiki/Exchange-traded_fund#Types AND http://etfdb.com/screener/.

[6] Robert L. Axtell. *Zipf distribution of U.S. firm sizes*. Science, 293:1818–1820, 2001.

[7] Eugene. F. Fama and R. French Kenneth. *Common Risk Factors in the Returns on Stocks and Bonds*. Journal of Financial Economics, 33:3–56, 1993.

[8] Y. Malevergne, V. Pisarenko, and D. Sornette. *Testing the Pareto against the lognormal distributions with the uniformly most powerful unbiased test applied to the distribution of cities*. Physical Review E, 83:036111, 2011.

[9] Didier Sornette. *Critical Phenomena in Natural Sciences (Chaos, Fractals, Self-organization and Disorder: Concepts and Tools)*. Springer Series in Synergetics, Heidelberg, 2nd ed., 2004.

[10] Jeremy Ramsden and Gy. Kiss-Haypl. *Company size distribution in different countries*. Physica A: Statistical Mechanics and its Applications, 277(1):220–227, 2000.

[11] Robert Axtell. *Firm sizes: facts, formulae, fables and fantasies*. *in* Claudio Cioffi-Revilla, ed.: Power Laws in the Social Sciences (Cambridge University Press), 2006.

[12] Herbert A. Simon and Charles P. Bonini. *The size distribution of business firms*. American Economic Review, 46:607–617, 1958.

[13] Orietta Marsili. *Technology and the Size Distribution of Firms: Evidence from Dutch Manufacturing*. Review of Industrial Organization, 27:303–328, 2005.

[14] Yuji Ijri and Herbert A. Simon. *Skew Distributions and Sizes of Business Firms*. Studies in Mathematical and Managerial Economics (Book 24), (North- Holland, Amsterdam), 1977.

[15] A. Saichev, Y. Malevergne, and D. Sornette. *Theory of Zipf's law and beyond*. Lecture Notes in Economics and Mathematical Systems (Springer), 632, 2009.

[16] Sandro Claudio Lera and Didier Sornette. *Effects of Mergers and Acquisitions on Firm Size Distributions*. Swiss Finance Institute Research Paper, (16-41), 2016.

[17] Y. Malevergne, A. Saichev, and D. Sornette. *Zipf's law and maximum sustainable growth*. Journal of Economic Dynamics and Control, 37(6):1195–1212, 2013.

[18] Jeffrey D Scargle. *Studies in astronomical time series analysis. II-Statistical aspects of spectral analysis of unevenly spaced data*. The Astrophysical Journal, 263:835–853, 1982.

[19] D. Sornette. *Discrete scale invariance and complex dimensions*. Physics Reports, 297(5):239–270 (extended version at http://xxx.lanl.gov/abs/cond–mat/9707012), 1998.

[20] Grzegorz Wilk and Zbigniew Wlodarczyk. *Tsallis Distribution Decorated with Log-Periodic Oscillation*. Entropy, 17(1):384–400, 2015.

[21] W.-X. Zhou and D. Sornette. *Statistical Significance of Periodicity and Log-Periodicity with Heavy-Tailed Correlated Noise*. Int. J. Mod. Phys. C, 13(2):137–170, 2002.

[22] W.-X. Zhou and D. Sornette. *Evidence of Intermittent Cascades from Discrete Hierarchical Dissipation in Turbulence*. Physica D, 165:94–125, 2002.

[23] W.-X. Zhou, D. Sornette, and V. Pisarenko. *New Evidence of Discrete Scale Invariance in the Energy Dissipation of Three-Dimensional Turbulence: Correlation Approach and Direct Spectral Detection*. Int. J. Mod. Phys. C, 14(4):459–470, 2003.

[24] Y. Huang, A. Johansen, M. W. Lee, H. Saleur, and D. Sornette. *Artifactual Log-Periodicity in Finite-Size Data: Relevance for Earthquake Aftershocks,*. J. Geophys. Res. (Solid Earth), 105:25451–25471, 2000.

[25] W.-X. Zhou, D. Sornette, R.A. Hill, and R.I.M. Dunbar. *Discrete Hierarchical Organization of Social Group Sizes*. Proc. Royal Soc. London, 272:439–444, 2005.

[26] Benedikt Fuchs, Didier Sornette, and Stefan Thurner. *Fractal multi-level organisation of human groups in a virtual world*. Scientific reports, 4, 2014.

[27] Wei-Xing Zhou and Didier Sornette. *Generalized q analysis of log-periodicity: Applications to critical ruptures*. Physical Review E, 66(4):046111, 2002.

[28] Wei-Xing Zhou and Didier Sornette. *Nonparametric analyses of log-periodic precursors to financial crashes*. International Journal of Modern Physics C, 14(08):1107–1125, 2003.

[29] S. Gluzman and D. Sornette. *Log-periodic route to fractal functions*. Phys. Rev. E, 65:036142, 2002.

[30] Didier Sornette. *Linear stochastic dynamics with nonlinear fractal properties*. Physica A: Statistical Mechanics and its Applications, 250(1):295–314, 1998.

[31] Per Jögi, Didier Sornette, and Michael Blank. *Fine structure and complex exponents in power-law distributions from random maps*. Physical Review E, 57(1):120, 1998.

[32] G Ouillon, D Sornette, A Genter, and C Castaing. *The imaginary part of rock jointing*. Journal de Physique I, 6(8):1127–1139, 1996.

[33] C Tsallis. *Possible generalization of Boltzmann-Gibbs statistics*. Journal of Statistical Physics, 52(1-2):479–487, jul 1988.

[34] Adrian Dragulescu and Victor M. Yakovenko. *Statistical mechanics of money*. Eur. Phys. J. B, 17:723, (2000).

[35] D. J. B. Soares, C. Tsallis, A. M. Mariz, and L. R. da Silva. *Preferential attachment growth model and nonextensive statistical mechanics*. EPL (Europhysics Letters), 70:70–76, April 2005.

[36] J. Cleymans and M. D. Azmi. *Large Transverse Momenta and Tsallis Thermodynamics*. J. Phys. Conf. Ser., 668(1):012050, 2016.

[37] Eugene F Fama and Kenneth R French. *The cross-section of expected stock returns*. the Journal of Finance, 47(2):427–465, 1992.

[38] Eugene. F. Fama and R. French Kenneth. *A Five-Factor Asset Pricing Model*. Journal of Financial Economics, 116:1–22, 2015.

[39] Thomas Piketty. *Capital in the Twenty-First Century*. Harvard University Press, 2014.

[40] Claudio Campanale. *Increasing returns to savings and wealth inequality*. Review of Economic Dynamics, 10:646–675, 2007.

Part II

Methodological contributions to temporal network theory

5

Dynamical properties of interaction data

5.1 Abstract

Network dynamics are typically presented as a time series of network properties captured at each period. The current approach examines the dynamical properties of transmission via novel measures on an integrated, temporally extended network representation of interaction data across time. Because it encodes time and interactions as network connections, static network measures can be applied to this "temporal web" to reveal features of the dynamics themselves. Here we provide the technical details and apply it to agent-based implementations of the well-known SEIR and SEIS epidemiological models.

5.2 Introduction

Network measures provide useful insight into the structure of relationships and interactions, and the breadth of systems that can be represented as a network has fostered an explosive growth in network analyses across all disciplines (for an introduction see [1]). Separately, analyzing system dynamics is common to all sciences; sometimes as differential equations, other times as regressions on time series data, and yet other times in animations or sequences of large patterns in data (e.g., changes in spatial maps or structural diagrams). But these standard

approaches capture only the dynamics of a measure rather than a true measure of the system's dynamics. In order to measure a system's processes directly, the methodology presented here captures the interaction and/or structural dynamics in a temporally extended network representation. Network measures applied to this "temporal web" then reveal features of the processes itself.

A number of recent papers have outlined similar techniques for capturing dynamic networks and/or dynamics on networks using a layered graph structure [2, 3]. Although these other temporal graphs have layers through time, those layers are not connected and the analysis focuses on time slices of the graph. Using the standard or adapted network measures for structural properties, values for time slices are calculated and compared to other slices. Our approach utilizes connections across time, i.e., spanning multiple slices. While standard "temporal networks are not graphs" [4], but rather series of graphs in temporal layers, our technique creates a single graph through transtemporal edges. We further differ in that our analytical technique is applied to this temporally extruded graph (i.e., the graph generated by connecting the layers) and its cross-temporal subgraphs rather than to time-slice layers. This makes a very different class of measures useful and provides distinct insights into the system's dynamics.

To demonstrate the technique we apply it to agent-based models (ABM) of both SEIR and SEIS epidemiological systems (i.e., a disease for which people go through the stages susceptible, exposed, infectious, and removed/recovered and alternatively in which people become susceptible again after the infectious state). This class of models offers a good test case because it has been well explored and is analogous to processes in many fields outside epidemiology (spread of ideas, technology, ...) [5]. Using an ABM instead of a categorical or stochastic differential equation model allows us to record the actual transmission events as well as perform agent- and time-specific contingency analyses. The purpose of this paper is not to contribute substantively to the modeling of SEIR or SEIS dynamics, but rather to demonstrate a new technique for measuring those dynamics. For this reason we have calibrated our SEIR ABM models to the ones presented by Rahmandad and Sterman [6]. By doing so we intend to inherit their description and minimize our exposition of the ABM model, its comparison to differential equations models, and the implications of heterogeneity in the ABM for health policy and transmission effects. Our SEIS model uses the same parameters as the SEIR model, and furthermore is build using the same interaction "skeleton" and random seeds as the SEIR models to provide the closest comparisons (more details below).

The purpose here is not to endorse a particular form of modeling or to contribute substantively to epidemiology in a direct way. The purpose is to present a *methodology for examining intertemporal interaction data*, and the agent-based modeling approach of this epidemiology problem is convenient for generating data with the appropriate structure. It also provides a clear demonstration that the

method produces additional insight into an already well understood phenomenon. The key is the interaction data. Upcoming research utilizes data collected from transactions among banks, neural activity (connectome) data across brain regions, inventory flows in logistic systems, and traffic patterns on road-rail-flight networks.

Techniques such as this one are important for delivering the potential of computer simulations of complex adaptive systems. For example, the cornerstone concept of *emergence* is purported to be a property of a system's interactions and dynamics, and therefore formal definitions of emergence may become available by analyzing the temporal webs of such systems. We do not attempt that task here, but instead make progress on current methodologies' inadequate ability to capture dynamics and analyze processes. For example, the projection or flattening of the interaction data into a single-layered graph makes the standard collection of network measures applicable, but it produces unreliable and/or incorrect propagation dynamics [7]. Our evaluation of the usefulness of existing measures (such as diameter, clustering, betweenness, eigenvector centrality, ...) on the temporal web revealed that these intuitive and useful time-slice measures fail to provide useful information on the temporally extended network representation (for reasons explained below). We therefore propose a new suite of measures grounded in network flow, in- and out-components, time-reversed in- and out-components, and transtemporal motifs. We demonstrate the effectiveness of this class of measures for identifying the agents *and times* upon which disease spread is the most contingent using a comparison to an exhaustive temporal knockout (TKO) measure.

5.3 The Agent-Based SEIR and SEIS Models

Our agent-based model of SEIR disease progression and spread follows the specification in Rahmandad and Sterman [6].[1] For the SEIS we used an identical setup and parameters, only changing the I→R transition to I→S.

Our simulations contain 200 agents that are connected in an explicit undirected base network – the collection of potential interaction conduits. The probability that agent i contacts agent j, given they're connected in the base network, is:

$$c_{ij} = \frac{\frac{1}{k_j}}{\Sigma_{K_i} \frac{1}{k_n}}$$

with k_j being the undirected degree of agent j, and the summation in the denominator is over each network neighbor (n) of node i (written K_i). For the results presented here we include only a fully connected *base* network which makes agent

[1] We provide the core details here for ease of reference, which will suffice for those already familiar with SEIR models. For more details about their ABM model, and their comparison to differential equations models, please refer to their paper.

interaction uniformly random over the full set of other agents. In our follow-up research we also analyze random, small world, and scale-free base networks following Rahmandad and Sterman [6], but we have omitted those scenarios in order to maintain focus on presenting the novel analysis technique being described in this paper. Updates are performed synchronously, so that that each agent's state at t depends on the states of agents at $t-1$. For each trial, these contact and state-changing dynamics are run for 400 time steps in order to ensure most SEIR infections are able to run their course and most SEIS infections reach full penetration.

Each individual has four key parameters: (1) expected contact rate, (2) infectivity, (3) emergence time, and (4) disease duration. The values for these characteristics (below) are calibrated to statistics for the common cold according to Rahmandad and Sterman [6]. Agents are in one of four states: susceptible, exposed, infectious, or recovered. Disease spread dynamics follow these rules:

- Initially two agents (1% of the population) are chosen uniformly at random and set to the infectious state.
- If an exposed agent contacts a susceptible agent, then the latter has a probability $i_{SE} = 0.05$ to become exposed (i.e., enter the first stage of the disease).
- If an infected agent contacts a susceptible agent, then the latter has a probability $i_{SI} = 0.06$ to become exposed.[2]
- At each time step, exposed agents have a probability of $1/15 = 0.066\bar{6}$ to enter the next stage of the disease, and become infectious.
- Infectious agents likewise have a probability of $1/15$ to become recovered in the SEIR case, or to become susceptible in the SEIS case.
- Recovered agents stay in that state indefinitely, and so they can also be thought of as removed.

5.4 Building a Temporal Web

The measures of dynamical properties utilized in this paper depend on a specific temporally extended network representation, so we now detail how to construct

[2] In this model the exposed state can be seen as an infectious but non-symptomatic state, whereas during the infectious state the agent shows symptoms. In many SEIR models the exposed agents are considered to be in an incubation stage and therefore not able to infect other agents (which matches the labels of the states better). In those models the time spent in the exposed state is referred to as a latency period between initial infection and overt illness. Thus the Rahmandad and Sterman model can be seen as an S-I_1-I_2-R model if one prefers that nomenclature. The distinction is not an important difference for our purposes here.

this representation. The "temporal web" presented here is distinct from other "temporal graphs", "temporal networks", or "layered networks" in its connections across time; however, they are all variations on a theme. As this subfield matures a better nomenclature will become necessary to disambiguate the techniques, although we will not propose one here.

The first step in building a temporal web is to decide upon a property to track: for us here it is the disease state. In our example we generate a temporal web directly from the model dynamics described above, but the technique is not a way to build generative models. Typically one will feed the appropriate interaction and state-change data (either empirical data or from a model's output) into a post-processing temporal web generating and analysis routine (for which we plan to offer a python package in the future).

In other applications, the original model or dataset may include a tremendous level of detail. Our SEIR system only has one property of interest, but separate temporal webs can be constructed from the same system for each property that one wishes to analyze in this way. These interactions may represent any sort of relationship among those elements: e.g., physical proximity, sharing of an idea, flow of resources, level of attraction, social obligations, financial debt, or any of the myriad relations that have been (or could be) encoded in a network. The nodes may likewise encode any properties that (potentially) change in response to what is chosen to be represented as the edges.

The only requirements for building a temporal web from a dataset is that it includes both:

- State changes in the elements across time and
- Interactions and/or relationships among the elements (e.g. agents) indexed by time.

Each period in the dataset becomes a layer in the temporal web; every element at each period is represented as a node in that layer.[3] The first period of state data is assigned to $t = 0$ and captures the initial values for the nodes. For each element, the node representing that element at t is connected to the node for the same element at period $t+1$ (if it exists) with a directed edge following the flow of time. These "temporal edges" connect the layers through the assumption that every element interacts with, or is related to, its former self. Naturally this assumption can be relaxed when appropriate.

Network connections among the elements (such as social connections, transactions, spatial relations, etc.) exist at the layer(s) representing when they occur. If the edges represent an existing *relationship* (directed or undirected) among the elements, then those relationships will usually be attached between pairs of nodes at

[3] An expansion of the technique under development allows for continuous time and heterogeneous time intervals (i.e., event-driven dynamics).

the *same time step* (intra-temporal). However, if the edges represent an *interaction, behavior, or process*, then that often happens across time and the nodes should be linked with a directed edge *across time* accordingly (cross-temporal). It is possible to have both flavors of inter-personal edges (intra-temporal and cross-temporal) in the same temporal web (e.g., ideas can spread immediately through talking and also by mail which takes time to arrive). For the synchronously updated SEIR and SEIS models used here, the inter-personal edges represent potentially disease-transmitting interactions that use agents' states at time $t-1$ to determine the states of other agents at time t, and so we will build a temporal web with cross-temporal inter-personal edges.[4]

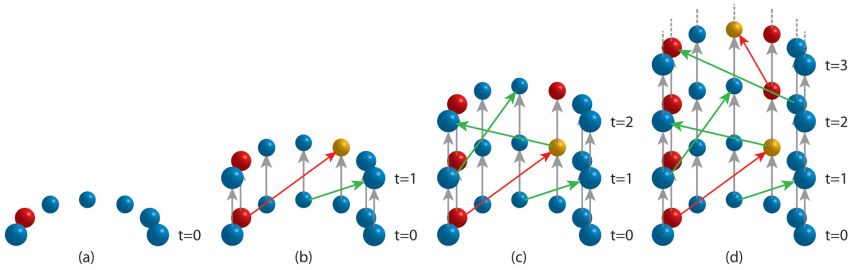


Figure 5.1: A simplified example of building the temporal web from state-change and interaction data for the SEIR model. Note that the interaction edges are cross-temporal to capture simultaneous updating in the generated data.

Figure 6.1 demonstrates the building process with just seven nodes (referred to as nodes 1-7 from left to right). In panel (a) the initial state of each agent is drawn for $t = 0$. According to the data used to build this temporal web, there were two interactions in the first period: from the infectious (red) agent 2 to agent 5, and from agent 4 to agent 7. Disease-transmitting edges are colored red while safe interactions are colored green. As a result of those transmissions agent 5 becomes exposed at $t = 1$ and is represented as yellow in (b). All other agents inherit their state from the previous period.

This procedure is continued to period $t = 2$ shown in panel (c). The exposed agent 5 communicates with node 1 but it does not transmit the disease. The data also reveals that between $t = 1$ and $t = 2$ agent 5 changed state from exposed (yellow) to infectious (red). All other agents are again unchanged.

From $t = 2$ to $t = 3$ (d) we again have two interactions: from the now infectious agent 5 that exposes agent 4 (red arrow), and from the susceptible agent 6 to the infectious agent 2. That this interaction *from a susceptible* agent *to an*

[4]Other research in progress applies this technique to empirical data (e.g., interbank loan data) that includes a detailed description of building a temporal web with temporally flat edges representing all transactions within a time-span. Though the analysis algorithms apply unchanged, there are difference in the constraints, caveats, and interpretations.

infectious agent failed to transmit the disease is not an assumption of the temporal web, that is simply read in from the data (i.e. it is either true of the empirical data or a product of the model used to generate this data).

The temporal web rendering process continues like this until all data points are represented. Larger sets of agents, longer time periods, more interactions, and heterogeneities in agent numbers and interaction lengths across time add *complication* to this procedure (and time to the analysis), but the rendering process follows the same steps as this simplified example.

5.5 Temporal Web Analysis

Visualizing the crystallized dynamics of your model/data in a static temporal web structure may provide some immediate insights that cannot be realized from animations or equilibrium states, such as patterns in activity or transient effects. Although useful for conceptualizing a system, those benefits are limited and qualitative. Because this way of generating a temporal web results in a single graph (rather than layers of graphs) existing measures from network and graph theory can be directly applied (with reinterpretation) to the temporally extended graph representation. And because we chose an SEIR/SEIS model as our example, there are also available measures from epidemiology that can be derived from the temporal web, indicating useful interpretations for other systems captured in this way. All that notwithstanding, a deeper understanding on the dynamics can be derived through a suite of analysis algorithms that have been adapted to harness the unique characteristics of this representation.

5.5.1 Standard SEIR Measures

Among the standard measures of disease virulence, prevalence, and morbidity the *reproduction number* (R_0) is the dominant metric. As succinctly stated in [6]: "A central parameter in epidemic models is the basic reproduction number, R_0, the expected number of new cases each contagious individual generates before removal, assuming all others are susceptible." They report that a differential equation model using the same parameters we used for the ABM yields an R_0 value of 4.125, making this a highly infectious disease.

Cumulative cases (CC) captures the number of individuals that ever become exposed or infected in the course of the disease. As a measure of impact it captures the breadth of infection and approximates the total morbidity. The *peak number* (PN) is the maximum number of agents sick at the same time. Because in the SEIR model the recovery rate is near the rate of new infections, the number of agents that are exposed or infected (EI agents) hovers in a fairly narrow band

compared to the cumulative number of cases.[5] For diseases in which R stands for recovery, the peak number approximates the maximal impact the disease has on the functioning of society – how many people are compromised on the worst day. The time step at which the peak number is reached is the *peak time* (PT) (if there are multiple time steps with a number of EI agents equaling the peak number, then it is the first of such periods).

The temporal network approach facilitates another intuitive measure of disease morbidity that combines the cumulative number of cases and the length of their illness. The *temporal magnitude* is the proportion of nodes in the temporal web that are exposed or infected (EI nodes). Because the nodes represent an agent at a time-step, the number of nodes (N) is equal to the number of agents (A) times the number of time-steps (T).[6] Temporal magnitude for our application therefore equals $\frac{N_{EI}}{N} = \frac{N_{EI}}{AT}$. For intuition purposes it may be helpful to note that when only normalized by the number of agents this quantity matches the area underneath a curve representing the percent of exposed and infected agents at each period.

Given the parameters and methods of the infection model used here, the length of the illness for each agent after first infection is determined by the probability of transitioning from E to I ($p(E \rightarrow I) = 1/15$), and the probability of transitioning from I to R/S ($p(I \rightarrow R) = p(I \rightarrow S) = 1/15$); these rates are constant and homogeneous across the population yielding an expected infection length of 30 time steps for each infection.[7] Because in this case the expected length of infection is homogeneous and known *a priori*, for the SEIR scenario the temporal magnitude is strongly correlated with the cumulative number of cases: the correlation coefficient is 0.9939.

However, for other applications in which either the cumulative cases or the average length are more complicated, temporal magnitude offers a simple measure of morbidity that is not reachable through combining other measures. One only has to consider SIS, SEIS, and SEIRS variations within epidemiology to find such cases because recurrent illnesses do not increase the cumulative case measure. In our SEIS experiments the correlation coefficient of the temporal magnitude with the cumulative number of cases drops to 0.9512. This value is still quite high, however, this reveals that within the first 400 periods of the disease outbreak new agent infections are (unsurprisingly) the dominant driver of increased disease morbidity. We found that it typically takes nearly the full 400 periods for everybody to

[5] Among all SEIR runs $\mu(PN) = 24.54$ and $\sigma^2 = 15.10$ compared to an average of 95.90 cumulative cases. Out of the 1000 runs, 272 runs produce an infection with fewer than 30 cumulative cases and we refer to these runs as "duds". Removing the duds from the analysis yields $\mu(PN) = 32.35$ and $\sigma^2 = 9.24$ out of an average of 129.42 cumulative cases.

[6] This holds as long as there is no agent entry or exit. It is possible to incorporate birth/death processes with the appropriate modifications to node counting.

[7] We observe a mean disease duration of 29.843 time steps in the SEIR model across all agents for 1000 runs. In some runs, there are agents still in the E or I state when the runs terminate at 400 periods, thus truncating their disease duration.

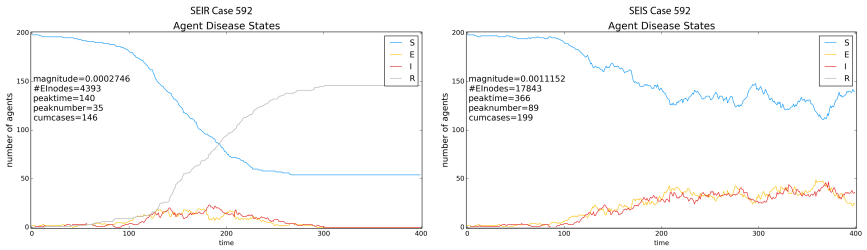


Figure 5.2: The progression of the disease states as well as 5 measures of disease morbidity for run 592 of the SEIR model (a) and the same run of the SEIS model (b). The choice of this case is explained in section 5.6.

become infected in the SEIS model given our interaction parameters. Thus we can expect cumulative cases and morbidity to become increasingly uncoupled as time goes on. Thus, in the contexts in which cumulative cases is appropriate, magnitude tracks it well; in the cases in which cumulative cases fails to capture morbidity due to reinfection, magnitude continues to measure morbidity.

Furthermore, when assessing the impact of quantitative agent states/properties across time (such as the debt of a bank, trade balance of nations, ...) the temporal magnitude measure generalizes to capture any attribute x of the agents (a_i). Summing a property's value across agents and across time will be informative only in certain contexts; for example, normalizing this by the number of nodes ($\frac{1}{AT}\sum_{t=1}^{T}\sum_{i=1}^{A} a_i(x_t) = \frac{1}{N}\sum_{n=1}^{N} x_n$) simply calculates the average property value across agents and time...a simple measure which has nonetheless demonstrated its usefulness time and time again. For the SEIR and SEIS models, the average property is equivalent to normalized magnitude when exposed and infected agents have value 1 and susceptible and recovered agents have value 0. This, we believe, is not a commonly presented measure of cumulative impact across time, yet it is an accurate representation of total morbidity and has clear applications to other systems (banks, nations, etc.). It can be used to compare the total morbidity of different scenarios, in a way that improves over cumulative cases and peak number. However, it cannot distinguish the different *timings* of disease instances, which we consider to be critical for understanding dynamical properties (and disease spread).

5.5.2 Temporal Knockout Analysis

In order to gauge the importance of each agent at each time for the spread of the disease, we propose a knockout sensitivity analysis inspired by knockout techniques used in genetic research [8] and ecology [9]. Temporal knockout (TKO) analysis determines how much a cross-time system property (such as total disease morbidity) changes when a particular agent is removed from that system at a par-

ticular time. That is, for each agent at each time, remove it and run the system process holding everything else constant to identify what effect that agent has on the system *and when*. The contingency of the system property to each agent-time's removal determines its *unique* causal influence on that property.

To achieve the ceteris paribus condition a preliminary step in performing a knockout analysis is to generate the *complete interaction dynamics* including who interacts with whom and which interactions would spread disease. We call this the temporal network "skeleton". For our models we generate an interaction structure based on the degree-depended interaction probabilities as well as store all the random numbers needed to govern the effects of interactions; i.e., whether each interaction would spread the disease if exposed or infected (which have different rates). We do this so that we can preserve the interaction structure and the contingent effects of each agent on the others while altering any of the infection parameters and disease states. Then for each *node* in the system, we remove it (set it to the recovered state) while keeping the interactions the same. The effect this removal has on the temporal magnitude of the disease is our measure of the *sensitivity of the dynamics to that agent at that time* (see Figure 5.3). We refer to this as the node's *temporal knockout score* or just *TKO* for short.

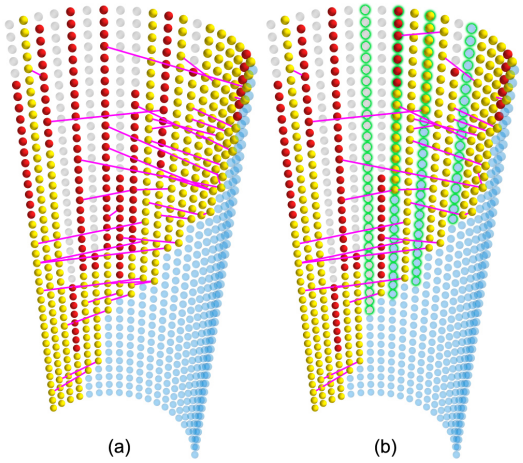


Figure 5.3: The difference in dynamics between the base case (a) and the case in which agent 7 is removed at $t = 10$ (b). The normalized difference in the number of EI nodes (i.e. the delta temporal magnitude) is that node's temporal knockout sensitivity score. Furthermore, there are variations in the exposed vs infectious profile (e.g. new positives generated vs infections averted) which can be picked up and differentiated by more fine-grained measures to produce multiple types of sensitivity.

The sensitivity of multiple system properties can be analyzed using temporal knockout, but we are focused only on total disease morbidity measured as temporal

magnitude. We believe that the TKO score is the best measure of an agent-time's influence on the system (at least for propagation models), and we therefore use it as the benchmark value against which all other sensitivity measures are compared. Because the knockout analysis requires rerunning the simulation for all periods after the node in question for each node in the temporal network, it is computationally expensive: $O(A^3T^2) = O(AN^2)$. Considering this needs to be done for every run of a model, it is prohibitively time consuming for large numbers of agents and/or long periods of simulated time. We therefore wish to find proxy measures that match the TKO rankings over nodes but with less computational time complexity. Clearly any proxy measure for contingent total magnitude will need to span both time and the agents, and we are exploring the possibility that network measures across the temporal web will fill this role.

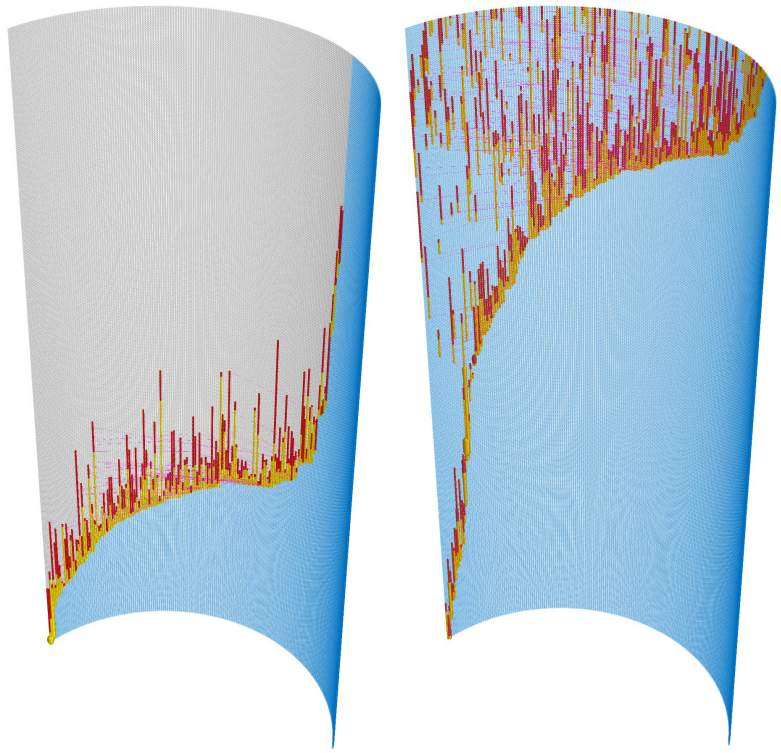
5.5.3 Standard Network Theory Measures

As described above, the directed edges connecting agents to their $t + 1$ selves produces a temporally extended single graph (rather than a temporally layered series of graphs). In this application we have also used cross-temporal interaction edges to capture the simultaneous updating of the model, but even when the interactions are restricted to their time-slices the temporal web is a single network (though possibly a disconnected one) . Because the temporal web structure is also a network/graph, some common off-the-shelf measures from network and graph theory can be applied to it. Naturally they must be reinterpreted to reflect the transtemporal meaning of the edges, and many common measures need to be adapted to work on directed graphs. We will see that most existing measures fail to measure interesting properties of the dynamics under this formalism.

It is also worth noting that while most measures on static graphs and temporally layered graph time-slices are agent-focused, our representation allows nodes that represent agent-times to be the focus of the analysis. Although we also want to tie it back to an agent for identification and intervention purposes, and eventually to compare the dynamical properties to properties of the agents themselves, as an analytical technique the shift from agents to agent-time nodes fosters distinct calculations and insights.

5.5.3.1 Out-Component

The out-component of a node is the set of nodes that are reachable from that node by following directed edges. The definition is identical to the static (directed) graph case, but now it tracks influence across time through cross-temporal links. In a temporal web the set of out-component nodes no longer represent specific agents, but rather all influenced agent-times of the focal agent. In our analysis here we can restrict the range of our analysis to the *EI-subgraph* of the interac-



(a) (b)

Figure 5.4: The full temporal web of one run (case 848) of the SEIR model (a) and one run (case 338) of the SEIS model (b). The yellow (E) and red (I) nodes together form the EI-subgraph representing the disease transmission in these scenarios. Although difficult to see at this scale, the nodes are sized by their OCPaths/EI score (see section 5.5.4.2) and these are cases in which that measure best matches the node-by-node TKO scores.

tion structure. The EI-subgraph consists of all nodes in state E or I together with the future-self connecting links between them and all the cross-temporal infecting links. Note that a cross-temporal infecting link can also connect to a node which already is in the E or I state. The number of EI-nodes in a node's out-component captures the proportion of the disease magnitude that the node could have generated. However, because there is typically a great deal of overlap in the out-components, this measure alone is not sufficient to capture unique, contingent contribution. By incorporating various refinements we use the out-component as the base for many of our temporal web measures presented below.

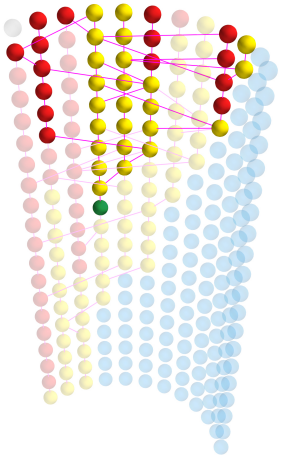


Figure 5.5: The out-component of a node (green) in the isolated EI-node subgraph. Note that in applications for which not all nodes are connected to their future selves (non-inheritance), the out-component may include a distinct set of individuals compared to temporally flattened graphs. Furthermore, in the SEIS model agents enter and leave and reenter the EI subset over time; as such, one's future self at a particular period may or may not be in one's own out-component.

5.5.3.2 Clustering Coefficient

Standard measures of the *clustering coefficient* report how frequently a node A is connected to a node C given that both are connected to a node B. The use of directed edges in this application already requires an alteration to the measure of the clustering coefficient, and cross-temporal interaction links makes true triangles impossible. Modifying this measure to track quadrangles of connection measures the frequency of interaction at time t given an interaction at time $t-1$, a property which would be useful for many models (e.g. as a measure of the amount of preferential interaction). Doing so is equivalent to measuring the edge-overlap of adjacent time-slices [2], which is a notion of clustering, but not akin to the original "friends of mine are also friends of each other" purpose of the clustering coefficient measure. For networks with interaction dynamics as sparse as our SEIR and SEIS models, the quadrangle measure is not (in practice) useful because the overlap of edges (in adjacent time-slices or otherwise) is frequently zero across the entire EI subgraph.

Capturing *temporal clustering* requires looking at larger inter-temporal patterns and determining how best to handle interaction directionality, time-induced directionality, and the span of time to consider. Given a timespan, and whether interactions are intra- or cross-temporal, one can determine the number of possible interaction patterns consistent with a triangle in the flattened subgraph. Each

such pattern constitutes an inter-temporal network motif equivalent to the "cluster" measured by the clustering coefficient. Figure 5.15 in section 5.7.2 on our motif detection extension illustrates one such example. A minimal cross-temporal example is: If A interacts with B and C in period t, do B and C interact with each other in $t + 1$? This has the same flavor of counting the number of triples that are triangles. The case of A interacting with B in t, and then B interacting with C and C interacting with A in $t+1$ also creates a $3t$-spanning triangle, but it doesn't have the "closing the triangle" look and feel. Most importantly, it will take considerably more work to determine if any of these triplet motifs provide useful information for temporal webs.

5.5.3.3 Redundancy and Efficiency

Redundancy and *efficiency* have technical definitions specific to network theory. Redundancy measures how many of a chosen node's neighbors are connected to other neighbors of the chosen node – for each node this is equivalent to one version of the clustering coefficient. The ratio of the redundancy to the number of the chosen node's neighbors is the network's efficiency. The lower the efficiency the greater the network density and number of redundant edges. One can also interpret redundancy as a knockout measure: what proportion of the network's "connections" (not just edges) are lost when that node is removed. As a measure of the robustness of network's connectivity it informs you of the number of nodes (or edges) that can be eliminated without compromising the propagation of the property. The weakness of these measures for temporal web applications is how the local node scores aggregate up to the global value. The kinds of reroutes that are implied by measuring redundancy on a flattened (or potential connection) graph are typically not possible on temporal webs.

However, the idea of measuring the robustness of propagation along paths can be measured in an alternative way for some temporal webs. Imagine an SI or SEI model – one in which nobody ever recovers (models of idea spread are often like this...no forgetting). In such a model we can trace the propagation *along every path* from the first infected to all the nodes infected at the final time step. Then for each agent-time node in the temporal web we determine what proportion of the paths are broken if that node is removed. Low scores indicate redundant paths and low efficiency. However, because the first nodes will always have extremely high scores, and the final nodes will have scores of zero, to measure a system's redundancy one will need to use the distribution of node scores to establish a redundancy/efficiency profile. The null hypothesis would be that redundancy scales linearly with the time progression, and deviations from that indicate particular properties of the system's dynamics. An approach similar to the Gini coefficient can then be used to compare redundancy across systems. The use of such a path-knockout calculation has other uses beyond pure propagation sensitivity, but because we do not deploy them for

the SEIR/SEIS analysis we leave deeper details to be described elsewhere. As you will see in section 5.5.4.2 we do use a notion of path redundancy to augment the out-component measure, but it is not used to calculate a measure like the standard redundancy measure.

5.5.3.4 Diameter

The *diameter* of a network is the longest geodesic path between all pairs of nodes; the minimal distance that ensures you can traverse the network. For a temporal web the diameter's constraints depends on the specific construction. If we put no restriction on the time-delay or time-span of interactions, then the diameter can be as small as 1 step when one of the initial nodes sends something to one of the final nodes. If the agent interactions/relationships are within the time-slices (i.e., layered graphs), then the smallest possible diameter is always $T + 1$ due to the essential directionality of transtemporal links. If inheritance isn't assumed for the layered time slices then the longest finite length that the geodesic path could be is $T(A - 1)$. When inheritance is assumed the worst-case scenario is length $T + (A - 1)$ because any agent that got revisited could just jump to its future self. If the interaction links are cross-temporal links, as they are in our SEIR/SEIS models here, then the diameter is always T (or infinity) because no path will use multiple edges per time step.

Determining the path length from each initially infected agent to every final agent across the EI nodes in an SI or SEI model reveals a combined measure of the time and the number of intermediaries necessary to reach the end state (or whichever time periods you choose). By subtracting T from the *maximum geodesic time-spanning path* one can find just the number of intermediaries required for a time-order-respecting diameter of the flattened graph (obviously only useful on layered temporal webs). In standard graphs, the diameter divided by the number of nodes is a useful, though basic comparison measure of overall network density/connectivity. Whenever inheritance is assumed the normalized diameter reports the same feature while respecting the time-ordering of links. When inheritance is not assumed, it reveals features of the interaction dynamics and connectivity that cannot be measured on flattened graphs. And because a temporal web is build from interaction data (with no explicit network required), the diameter can be used to analyze differences in connectivity (e.g. small world effects) found in a dataset's or an ABM's de facto interactions.

5.5.3.5 Closeness Centrality

The standard *directed closeness centrality* of a node is calculated as the inverse of the sum of distances to all nodes in its out-component (or the harmonic mean to all nodes). Because the graph is directed across time, it is never strongly connected,

and therefore closeness centrality is intuitively of little use. Suspending the directionality yields a measure of which node is most central to all other nodes in both receiving and sending influence, which may have some useful applications for information propagation. The result is not a measure of influence because it traces backwards in time, but it can act as a measure of how "in the thick of it" an agent is at at time step. Our measure of *nexus centrality* in section 5.5.4.6 is similar to this bi-directional closeness centrality.

Note that when interactions are cross-temporal the length of any path connecting two nodes is the time difference between them, which simplifies the calculation algorithm. In our SEIR/SEIS application the temporal edges dominate the temporal web's and the EI-subgraph's connectivity, so the later the period the closer the node is to all the nodes in its out-component. In other applications, temporally adapted variations of closeness can be used to rank the importance of both agents and time periods for achieving a specific end-state by only measuring distances to the target time period(s). For example, in the EI-subgraph of the SEIR model there are many dead-ends for the disease transmission. Tracking distance only along the infecting paths yields a measure such that greater closeness implies less long-term impact. The ratio of closeness to the out-component size indicates the balance of multiple short paths versus a few long paths because the measures are cumulative across all nodes along the paths. This and other versions are explored in future work.

5.5.3.6 Betweenness Centrality

Ordinary *betweenness centrality* is a measure of network structure that calculates for each node i the number of geodesic paths between all pairs of nodes that include node i. Among other things, it indicates bottlenecks in the connectivity between communities of nodes [10]. For the spread of a disease we are very much interested in bottlenecks of the sort that betweenness reveals in static connectivity graphs. In a temporal web analysis, we would like to utilize bottleneck detection to uncover crucial junctures in the way interaction dynamics unfold.

We applied the standard algorithm [11] and a directed-edge variation to each node in the EI-subgraph to determine each node's betweenness centrality in the system's infection dynamics. Because the temporally directed paths only flow in one direction we believed that nodes near the end of the time span will typically have the most paths going through them as they accumulated across time. However, what we found is that because betweenness only counts the *shortest* paths, the highest betweenness nodes occur in the busiest time steps (when the most inter-agent connections are found). In our experiments nodes are usually not connected by multiple time-respecting paths in the EI-subgraph. And when they are, the path length between nodes is always just the difference in time, so they are all the same length. As a result, betweenness scores are the same or very similar across

the nodes of an EI-subgraph in the current application, making them predictive of nothing.

Betweenness may be more useful for other applications, and it may be useful to make some alterations to its calculation for the temporal web domain. For example, using layered interactions will produce longer and varied path lengths, giving betweenness something to track. Going further, instead of calculating betweenness using the number of geodesic paths from all nodes to all nodes, we can consider a subsets of nodes; specifically, from one time period (e.g., $t = 0$) to another period (e.g., $t = T$). This *timespanning betweenness centrality* may reveal how much *and when* each element contributed to the transfer of the tracked property among the agents and across time.

We can refine our use of timespanning betweenness even further. We can relax our constraint to the EI-subgraph to calculate the timespanning betweenness of the full graph to identify *potential* super-spreaders rather than actual spreaders. This still differs from "betweenness preference" and other time-preserving betweenness calculations [7] by focusing on the propagation of properties across agent-times, but it becomes very close. The set of nodes identified with the greatest timespanning betweenness value may correspond to what many refer to as a *tipping point* or *critical transition* in the dynamics [12], which should correspond to the TKO score in disease propagation. That is, the identified nodes would mark an agent and time window in which removal (e.g. quarantine or previous inoculation) would be most effective. Future work exploring the effect of various underlaying network structures will determine whether this or other variations of betweenness can capture bottlenecks in temporal webs.

5.5.3.7 K-Core

The k-core measure has been increasingly used as a measure of propagation sensitivity in networks. For example, Karas and Schoors show that the k-core is most highly correlated with risk propagation in the time-flattened Russian interbank loan network [13]. It is also used in neural network analysis, idea (meme) spreading, metabolic networks, etc. The sparse interaction among EI agents in our model makes it so that for most nodes the only edges are the temporal edges connecting agents across time. However, those agents who undergo a short period of elevated interaction will have a large k-core robustness. Although k-core does not perform well in tracking temporal knockout, it might be tracking some other dynamical property very well. Future research will investigate what feature of interaction data, if any, the k-core tracks and where it may be uniquely useful. We are also working on an adaptation of the algorithm to find the KT-core – a time-integrating iterative k-shell removal technique to identify burst of activity in the temporal web.

5.5.3.8 Degree Centrality

We capture the out-degree centrality of every node in the EI-subgraph with the idea that a node's individual contribution to disease spread may be best captured by its direct influence. We also include the in-degree centrality to account for others' influence on it. These two combined are measure of a node's throughput, which could indicate that a node is important for how the dynamics unfold. Degree centrality performs well on temporally flattened graphs as a rough measure of influence, but on the temporally extended graph all our variations fails to capture the infection sensitivity. The reason is that all nodes have the out-edge to their $t + 1$ self and a few have one interactions, and a very few nodes have more than one interaction, and so every spreading agent will have the same out-degree. Node degree fails because it only captures the immediate effects, and the kind of contingency that parallels temporal knockout requires incorporating downstream effects as well. In section 5.5.4.1 we present extensions of degree centrality that further incorporate temporality through accumulating degree values. Although node degree does not capture the property we are looking for, it is worth thinking further about what it might capture in other models and other constructions.

5.5.3.9 Network Flow

Network flow is less of a measure and more of a method to calculate measures. There are many ways to construct and use a network flow algorithm, and many of our adapted measures presented below have an inspiration and calculation via network flow-type algorithms. Determining critical agents with network flow is already common in networked epidemiological models [5, 14], and it can also be used for community structure identification and robustness. As with other measures the temporal directionality of our construct undercuts much of the power of network flow approaches. With no cycles, and therefore no feedback, the standard measure is merely tracing the out-component, but it can be used in various alternative ways.

For example, if we start flow from each node, let it follow all out-going edges, and add the flow at each node to account for all nodes that the flow passes through, then we can use the stored flow as a measure of message reception: how much of the system's flow passes through each node. Looking at the full temporal web skeleton, we can use flow to determine the likelihood of nodes becoming infected for scenarios in which we don't know which agents will be initially infected. Essentially this determines the number of time-ordered paths that a node is on as a measure of infectability.

But this is not a metric that is useful for disease propagation in which influence over the future is what is important. For this application we can calculate the *time-reversed network flow* which scores previous nodes by their responsibility as

a source of infection for each later node. Or we can use it as another time-spanning technique that traces back to the actual or potential initial source for each unit of infection present at time $t = T$. Such time-spanning measures are good for models without recovery, but fail to capture dead ends in SEIR or SEIS disease spread; and dead ends in the disease propagation still contribute to overall morbidity. Hence network flow, normal and time reversed, is not a reliable proxy measure for temporal knockout here, but it is explored in depth in other work regarding the spread of technology and of bank risk leading up to a systemic collapse.

5.5.4 Temporal Network Measures

Graphs are tightly constrained representations that have been explored thoroughly over the past several decades. As a result there are few possible unexplored, simple measures of properties of networks. However, variations in the structure of graphs (such as k-partite graphs, weighted multi-graphs, hypergraphs, and layered graphs) open up new avenues of analysis. A temporal web as constructed here, however, is just a normal directed graph and therefore the measures that follow make use of familiar network property calculations. Below we present a few measures devised to extract information about dynamical properties from the temporal web to demonstrate the sorts of mathematical gymnastics that make use of the temporal interpretation of the directionality.

5.5.4.1 Cumulative Degree Measures

Cumulative in- and out-degree is the amount of a system property that feeds into and flows out from an agent, accumulated across time. For the SEIR model we can consider the property as 0 for the S and R states (not infected) or 1 for the E or I states (infected).[8] Summing in-degree for an agent going forward in time calculates how many times the agent has been infected up to each node's time-slice; we call this *uptake*. The cumulative out-degree, or *discharge*, measures how many times the agent has already infected others up to that time slice. The values of these measures at $t = T$ are the same as the respective flattened graph measures, and insofar as it tracks influence, the temporal web versions foster distinctions for when an agent's influence arises.

Running backward in time, the reverse cumulative in-degree, or *popularity*, measure reveals how many times the agent will receive an interaction in the future. The popularity measure could be useful for assessing the impact of vaccination schemes by informing us of both who and when a vaccination will protect an agent the most from likely infections. Finally the time-reversed cumulative out-degree

[8] In a more nuanced application we could choose a value for the exposed states between 0 and 1 that captures the relative rate of infection spread for those nodes. In general, it is not difficult to adapt these measures for real-valued node properties and/or edge weights.

measures how many times the agent will spread the property (i.e., infect) to others in the future. We can call this *gregariousness* which, like popularity, highlights the dispositional nature of measures over future behavior; these describe inputs and outputs if nothing intervenes.

These four variations on the degree measure can be combined to capture a variety of dynamical properties. Including the temporal aspect through cumulative sums allows these measures to go beyond immediate effects. Intuitively one could combine the information regarding how many more times an agent will receive an infecting connection and how many agents it will contact in the future to determine the benefit of vaccinating that person at that time; e.g., by multiplying popularity and gregariousness. However, due to the actual propagation patterns of diseases with our duration and infection parameters there are rarely more than a few interactions across time for any agent. In our models an individual's impact on overall morbidity occurs through the indirect downstream infections, which are not revealed through degree measurements. For other uses of temporal webs, such as the interbank loan network, in which interactions themselves are important indicators of system health, these temporal web measures of cumulative degree may reveal themselves to be more useful.

5.5.4.2 Out-Component Paths (OCP)

The standard out-component (described in 5.5.3.1) counts every node that is "downstream" of the focal node; i.e., every infected agent-time that this node may be responsible for (figure 5.6). However, it turns out that this, by itself, fails to account for several other features that are important for capturing temporal knockout sensitivity. The feature of betweenness that we wish to capture in a dynamical measure is its ability to identify bottlenecks in the spread of the disease. The out-component actually under-represents the out-flow from a node because there are multiple paths from the focal node to many of the nodes downstream. In this case, we are not particularly interested in the shortest path connecting them – any path will succeed in spreading the disease. Thus we sum the number of paths from the focal node to each node in its out component.

By counting the number of paths to a node we can better account for a node's potential, or contingent, impact on future morbidity. This allows a better match with the sensitivity analysis, but it is only a measure of output. To really capture the bottlenecks we want to measure not just high throughput, but high throughput with few alternative paths capable of maintaining that throughput. The problem with out-component paths is that there may be multiple agents with the same number of out-component paths (and many of the same paths) at a given time-slice. Although each one can be considered responsible for spreading the disease, none of them can be seen as crucial for spreading the disease. In order to match the sensitivities of the knockout analysis we further divide a node's out-component paths

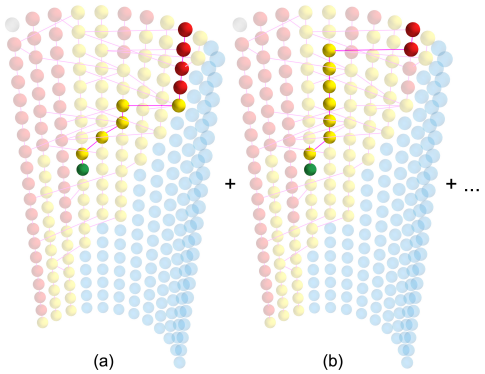


Figure 5.6: To calculate the out-component paths score for the green node, we sum the lengths of all the paths from this node to each of the nodes in its out-component.

(OCP) score by the number of EI nodes that exist at its time-slice (OCP/EI). This is still rather crude, and a technique that more carefully identified and adjusted for influence overlap may perform better, however we leave that analysis for future work because our results here show that it would be difficult to identify any measure that outperforms OCP/EI on these models.

Our algorithm for identifying all the paths for each node pair exploits the time directed nature of the temporal network structure. We start with the nodes at the last time period (those with no out-edges). Nodes at time T trace back across each incoming edge and adds itself to each $T-1$ node's list of out-component nodes. The $T-1$ nodes add themselves to their list of out-component nodes as they pass the list down to the $T-2$ nodes...and so on until $T=0$. The time-reversed path tracing allows us to calculate all possible pairwise paths while crossing each edge of the temporal web exactly once. For the case in which interactions are restricted to the immediate successor time period (such as our models here) the worst case computational time complexity for determining the out-component paths is $O(A^2T) = O(AN)$. Each node would need to add itself to the list of every node at the previous period and this would be required for every period.

5.5.4.3 Out-Component Paths Future In-Agents Weighted (OCPFIAw)

As we will see in section 5.6, the OCP/EI actually performs the best of the measures we developed, and better than any of the off-the-shelf network measures applied to the EI-subgraph. We also developed three other refinements that can be applied to the out-component or out-component paths that performed nearly as well in this application and may perform better in other applications. These refinements also demonstrate the kinds of features one can look for in the dynamics captured by temporal webs, and hopefully inspire others to develop their own

refinements.

The first of these refinements weighs each EI node by the number of *unique agents* with an edge to that node *or any future self of that agent*. This measure connects the temporal web measures with a measure more similar to the temporally flattened ones. It utilizes unique agents and includes all future connections to the focal *agent* (rather than just that node). Our reasoning is that this node, if not infected by the path currently being evaluated, would still become infected via one of these future in-agents (figure 5.7). Thus the contribution that this node adds to each of the paths using it is discounted by multiplying $1/(1+\#FIA)$; with *#FIA* being the number of future infecting agents.

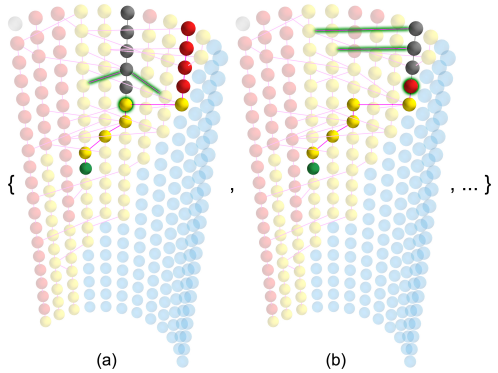


Figure 5.7: Each node's contribution to the out-component path measure is weighted by the number of incoming edges (green glow) from other agents that occur in the future of this node's agent (gray nodes). The greater the number of future in-edges from other EI nodes the less important this agent at this time is to the spread of the disease.

In a variant of this we also calculated a weighting based on *future in-nodes*, instead of agents (OCPFINw). The idea here is to capture the fact that those other agents have multiple infection paths running through them, and therefore counting it only once may fail to pick-up the disease contingencies properly. In practice, our SEIR-subgraph is sparse enough that the difference between in-agent and in-node weights is negligible. However, we mention it here both for completeness and because this adjustment (or a similar adjustment) may be precisely what is needed in other applications.

5.5.4.4 Out-Component Paths In-Component Weighted (OCPICw)

Another approach to weighing the OCP nodes adjusts for the total level of influence received by that node. For each node in the EI-subgraph we determine the size of its temporal in-component (all the *nodes* that have a path to it, IC). We then calculate each node's weight as $1/(1 + |IC|)$. Nodes with larger in-components

add less weight to a node's temporal centrality because it receives influence from a larger base and therefore contributes less to importance of the focal node (figure 5.8).

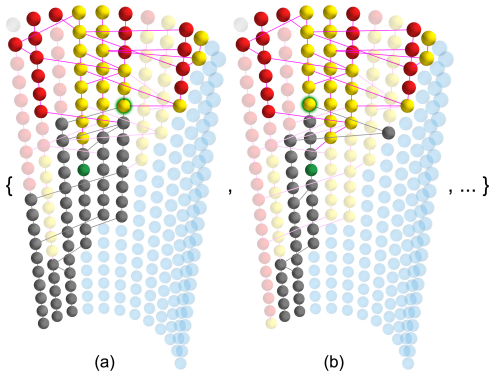


Figure 5.8: Each node in the EI-subgraph is assigned a weight based on the size of its in-component with larger in-components decreasing the weight of the node. The OCP measure is then calculated as the sum of these node weights across all paths connecting a focal node (green) to every node in its out-component.

Just as we have replaced the temporal out-component with the number of out-component paths, one could also calculate a modified version of this measure using in-component paths. A further refinement weighs a node's contribution to a focal node's score by the proportion of in-paths to that node on which the focal node lays. That is, if the focal node is on every path to the node, then it gets all its weight. If several other nodes can reach it, then the focal node's importance is diminished. Naturally there are further refinements to taking a node's accumulated receiving influence into account when assigning its contribution to other nodes' scores.

5.5.4.5 Out-Component Paths Redundant Paths Weighted (OCPRPw)

Our most complicated measure of temporal web centrality that we devise and present here, establishes a tradeoff between immediacy and redundancy. We first determine all the paths from the focal node to each node in its temporal out-component. Now we use the lengths of those paths and assign a weight to each node based on the sum of the lengths of the paths on which it lies (figure 5.9). Specifically, each node's contribution is the *number of paths it is on divided by the sum of the path lengths*. In this way, nodes that lie on short paths count more than nodes on long paths. If there are enough short paths then it can still equal the weight of a long path, but the idea is to discount nodes that contribute mostly to dead ends.

This measure, and other variations similar to it, require a certain structure of

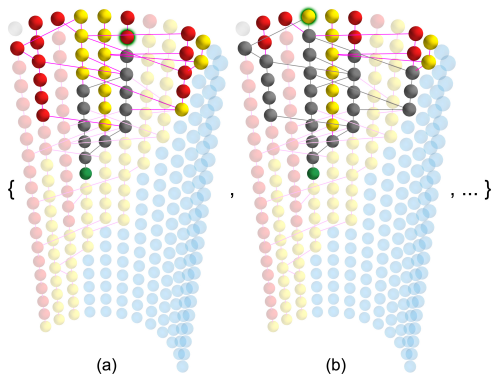


Figure 5.9: For each node in the focal node's (green) out-component we determine all paths between them and the lengths of those paths. This measure sums the number of node occurrences along all paths divided by the sum of the lengths of all paths each node is on. This complicated measure produces a tradeoff that weighs short and/or unique paths more and long and/or redundant paths less.

the data to be useful. For example, in this case there must be multiple paths of different lengths in the EI-subgraph in order for this measure to differentiate nodes by them. What we observe for the SEIR model is a sparse interaction graph with few alternative paths as well as disease impact dominated by the few initially infected. In our SEIS models the initial agents are still the most critical, although bottlenecks can exist. These refinements will be more useful for models with denser interaction networks, more paths of different lengths, and longer running times, such as in our interbank loan network application.

5.5.4.6 Nexus Centrality

We also introduce a temporal measure related to closeness centrality and betweenness that is calculated from components and can be weighted by any of the adjustments just presented. For each node in the temporal web (or in the appropriate subgraph) we determine all the paths in both the out-component and the in-component. Then for each node we multiply (1) the sum of the lengths of all paths to all nodes in the focal node's out component (OCP from above) and (2) the sum of the lengths of all paths in the focal node's in-component (ICP). This is the sum of the lengths of all paths running through the focal node. As with the other centrality measures this can be utilized in various ways (e.g., using a subgraph of "infected" nodes or using a time-spanning measure from $t = 0$ to $t = T$) as appropriate to the data and the information desired. Because the in-component is not an important feature of disease spread we omit this measure in the results below, but preliminary results for identifying bottlenecks in ongoing processes (such as debt

risk) are promising.

5.6 Results

We ran both the SEIR and SEIS agent-based model 1000 times and collected a full battery of measures: standard SEIR measures, standard network measures on the EI subnetwork, and our temporal web measures with various alternatives – 79 measures total. Recall that, as distinct from analyzing time-slices, we want to analyze the dynamics themselves by applying a measure to the whole temporal graph and identify dynamical properties. Specifically we wish to identify those agents with the greatest impact on the disease spread and when they are key agents. We do this by comparing the scores for each node for each agent against its Temporal Knockout (TKO) score. As we explained in section 5.5.2, the TKO score provides the best contingency sensitivity test because it simulates the same infection dynamics considering the elimination of each node and measures the resulting change in disease magnitude.

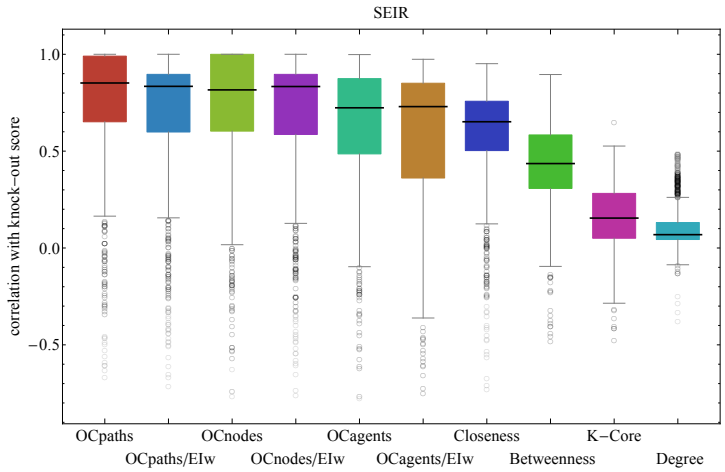


Figure 5.10: The summary results from 1000 simulations of the SEIR model showing the node-by-node Pearson correlation between the temporal knockout score and several temporal web measures and off-the-shelf network measures.

Both SEIR and SEIS dynamics create a situation in which the initial agents typically have the largest impact on downstream infections. This is not always the case, but when this is the case it is no surprise that out-component based measures will accurately capture that dependency. In our SEIR experiments 23.5% of the runs infect fewer than 10 people (using the cumulative cases measures) and 27.0% yield fewer than 30 individual infections (which we call “duds”). Our SEIS ex-

periments (which use the same random seeds) similarly produce 23.2% with fewer than 10 and 26.5% with fewer than 30 individuals. Our temporal web measures typically perform extremely well in these dud runs (clusters of points on the left of figure 5.12) which may be upwardly biasing our performance results on more important scenarios. Specifically, we are interested in the the temporal web measures because they should be able to pick up properties of dynamics, but the properties we are interested in cannot exist in runs with so few infections. We performed the measure comparison analysis with and without the dud runs, and the difference was too small to change the order of the measures' performance and therefore we present only the results with duds included.

Due to the large scale of our experiments and the number of variables we measured we provide only a summary of the overall results along with a detailed analysis of our best measure to provide the insight into how temporal web measures differ from traditional network measures and what they reveal. Figure 5.10 reports the correlations of ten measures across the 1000 runs of the SEIR model. OCpaths and OCnodes perform nearly the same in both the standard and divided-by-EI-nodes-this-timeslice (/EIw) versions. OCagents, which is equivalent to the standard out-component on the time-flattened network, performs only slightly worse. The off-the-shelf network measures all perform worse on average when applied to the temporal web, although closeness and betweenness do match TKO well on a number of runs. For reasons explained in the network measure descriptions above, the standard K-core and degree centrality measures are poorly adapted to temporal web applications with cross-temporal interactions and that is revealed clearly in their poor performance here. Closeness does as well as it does on the SEIR model because it assigns the nodes at the earliest time periods with the greatest closeness score; thus it is picking up a superficial property of SEIR dynamics on a time-directed graph rather than an important, general property of propagation dynamics.

As compared to SEIR models, the SEIS model has a much greater capacity for disease morbidity both in terms of the number of agent-times that can be in the EI state and the probability of infection through an interaction. As time goes on in the SEIR model the pool of possible infectees decreases because more and more agents are immune, whereas an SEIS model on the same skeleton will use the same edge as a re-infection. You can see the differences in magnitude in figure 5.12 in which the blue marks indicate SEIR results and red marks indicate SEIS results. Greater magnitude mean greater contingency, and reinfection means that latter periods are more likely to be important. Although in the growth stages of a disease, such as we modeled, the initial agents are still most likely to be the most important, the potential for a single infected person late in the process to generate a large magnitude is much greater. Thus matching the TKO in an SEIS is more difficult.

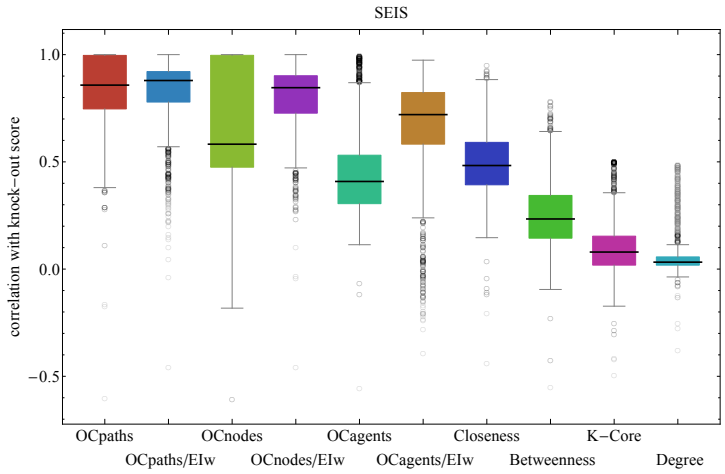


Figure 5.11: The results from 1000 simulations of the SEIS model showing the node-by-node correlation between the temporal knockout score and several temporal web measures of dynamical properties.

The results of our SEIS simulations are presented in figure 5.11. Here we notice that dividing by the number of EI nodes in a time slices drastically increases the performance of the temporal web measures in the SEIS model, although it had no consistent effect in the SEIR model. Dividing by the number of EI nodes is intended to factor in the number of alternative paths of disease spread, and thus how bottleneck-like a node could be. SEIR dynamics largely eliminate the potential for bottlenecks, but they are important for SEIS infection dynamics, and this difference is starkly revealed in the difference in performance of the "$/EI$" measures here.

Also, as was expected, the OCagents measures (which are identical to the flattened network out-component) perform worse on the SEIS models because they cannot track increased morbidity from reinfection. Closeness performs much worse than in the SEIR model precisely because early agents are less important in SEIS dynamics. The other standard network measures perform consistently worse on the SEIS model as well. OCpaths and OCpaths/EI both perform *better* on the SEIS model, although our various other weightings on the paths did not make any further improvements.

The ability of OCpaths to track the greatest TKO agents so well can be understood through the features of the measure. It combines the rough measure of total impact of the out-component with the ability to count the multiply infected and incorporate a measure of redundancy by counting the nodes along each path. It can also be quickly calculated, making it a desirable proxy measure for the com-

plete knock-out analysis. By examining the performance of the measure in more detail we may be able to identify those features of the dynamics that effect the performance the most, and thus further improve our ability to capture the dynamical properties we are interested in.

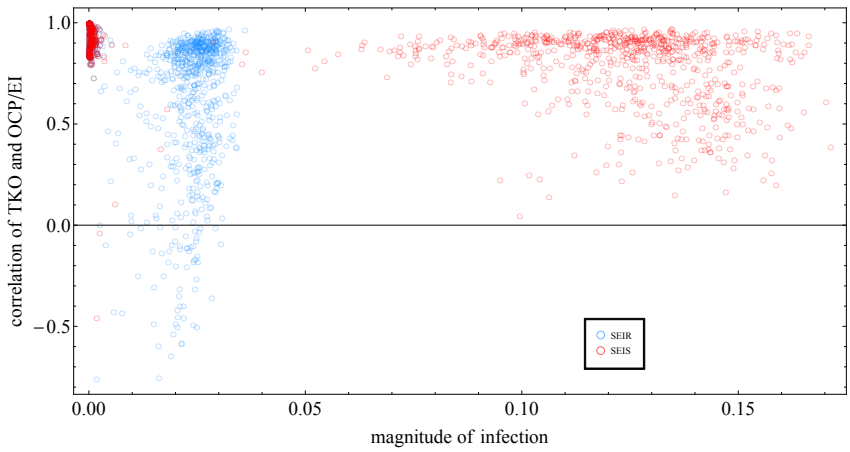


Figure 5.12: This scatterplot shows the relationship between the morbidity of the disease as measured by the proportional size of the EI subgraph (temporal magnitude) and the correlation of OCP/EI and TKO. When the diseases fail to spread we can easily identify the key, early individuals. As the disease spreads more the performance becomes highly variable, and in some SEIR runs the OCP/EI measure is anti-correlated with TKO although it performs well overall in both scenarios.

Figure 5.12 shows the relationship of the magnitude of infection and the performance of OCP/EI measured in terms of Pearson correlation with TKO. Figure 5.13 instead compares the correlation with the fraction of periods in which the OCP/EI measure accurately identified the highest TKO individual. This latter performance measure is meant to reflect the usefulness of the temporal web measures for selecting the best target for intervention policies. If multiple agents could be vaccinated, quarantined, or otherwise removed from the system, an overall high correlation between OCP/EI and disease sensitivity indicates that our measure is generally a reliable proxy, and better in SEIS scenarios than SEIR scenarios. Comparing the two figures shows us that high correlation almost always parallels matching the top agent each period, and overall lower correlation is not strongly tied to the magnitude of the infections. Specifically, once one removes the duds, a linear regression reveals no significant positive or negative relationship between magnitude and correlation whereas the matching fraction and the TKO correlation are strongly positively correlated for both the SEIR and SEIS data.

For our simulations the overall fit, rather than only matching the highest nodes,

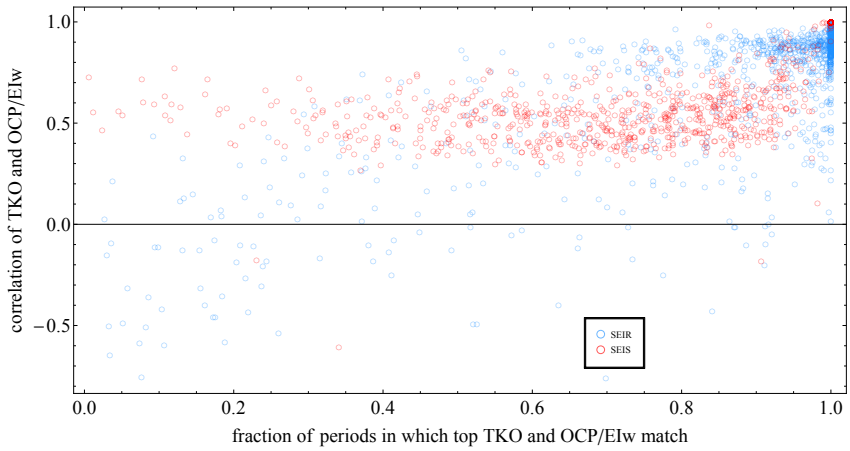


Figure 5.13: This scatterplot shows the relationship between identifying the highest TKO agent (x-axis) and the node-by-node correlation with the TKO value (y-axis). The variation in correlation cannot be sufficiently explained by matching just the most sensitive nodes, thus indicating an overall good fit.

was important to determine because we found that the gap between the highest and second highest score per timeslice was often quite large. The property of carrying the greatest TKO and OCP value appears to move from agent to agent across interactions as time progresses, only rarely are there multiple agents with high values at the same time step. Thus it seemed possible that matching only this highest agent could explain most of the correlation, our investigation into the fraction of max-TKO matching shows that this is not the case. Furthermore we did not find and consistent patterns in the relationship between max-TKO levels (samples shown in figure 5.14) and overall measure performance.

The four cases in figure 5.14 show the best-case and worst-case scenarios for OCP and OCP/EI in both the SEIR and SEIS models. Actually there are many cases in which the OCP and OCPDIVEI measures match perfectly, so I have presented here the cases with the greatest temporal magnitude for each one. Case 592 is conspicuously the worst-case for both models. Recall that both models use the same skeleton; i.e.; the same random seed and the same sets of interactions. The *only* difference between these two scenarios for case 592 is whether agents become susceptible or recovered. And for both models this case produces nodes with negative TKO values. When a node has a negative TKO score this means removing that agent at that time *increases* the total disease spread. This can happen when the interaction structure is arranged in a way such that if some particular agents are infected early on they quickly recover without spreading the disease, but the same agent infected later will stay infected and infect many others. Identifying

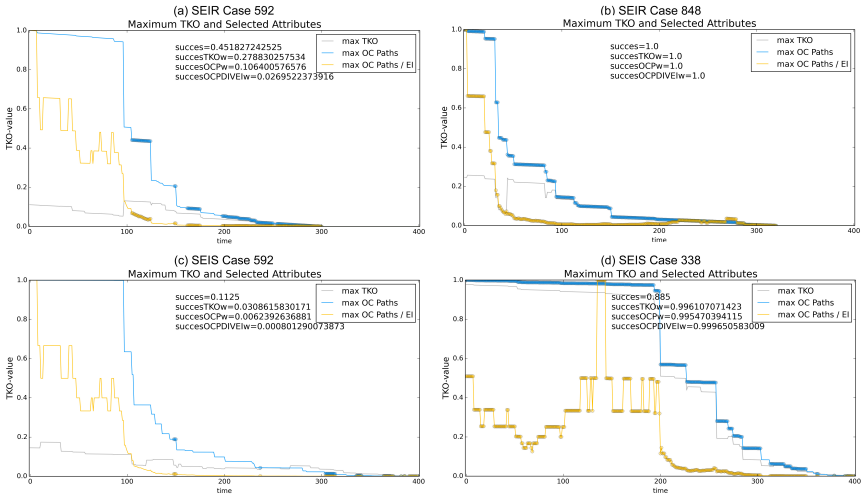


Figure 5.14: This plot shows the success of OCP/EI, our overall best measure, to identify the highest TKO agent at each time step as well as demonstrate some variations in the TKO patterns. The embedded scores indicate (from top to bottom) the raw proportion of OCP correct matches, the OCP match proportionally weighted by the TKO value, the OCP proportion weighted by the OCP value, and the OCP/EI weighted by the OCPDIVEI value. Time steps with disc marks indicate a match between OCP and max-TKO.

them is tricky because it is not the focal agent being knocked out, but rather one of the agents that the knocked-out agent infects. That is, a node will have a negative TKO if at least one of the agents it infects (directly or indirectly) would produce greater disease morbidity if it were instead infected at a later time. And not just greater than that too-soon-infected agent created, the increase must be greater than all the magnitude removed by knocking out the focal node. The reality of this counterintuitive possibility is partly responsible for the low performance of case 592, and it also points to interesting feature of SEIR/SEIS dynamics still in need of accurate measure.

Our results show that despite the less-than-perfect match overall and high variation across some runs, our out-component paths and related measures do outperform standard network centrality measures on the EI-subgraph. Recall that this is not a comparison with these network measures applied to the temporally flattened graph equivalent of the dynamics (that is coming in future work). We applied these off-the-shelf measures to the time-directed EI-subgraph of the dynamics. The directed edges of this network, as well as the sparseness of interactions across time, account for the poor performance of the standard measures in this context. Nascent techniques for time-layered networks cannot be applied here because there are no layers in version of a temporal web with cross-time interactions.

Path-finding and network flow approaches to measuring dynamical properties

seem to offer the best results, however the field of temporal networks is still young and better measures may be found as it matures. What is probably more important to keep in mind is that these measures performed better *at the task of tracking disease impact across random interactions*. Changing the task will shift which dynamical properties are relevant, and that is going to change which measure is best at capturing the desired property. It is for this reason that we presented not just our best measure and the SEIR/SEIS results, but also other varieties of measures and our thinking behind them. Researchers interested in other properties of other models are invited to use, adapt, improve, and reinterpret the measures provided here. The SEIR/SEIS results are indicative of the benefits of temporal web analyses, gaining additional insight into such well-tread territory.

5.7 Extensions and Modifications

As already mentioned, this paper presents the temporal web measures with interpretations appropriate for the SEIR and SEIS models. Applying the temporal web technique to other models, data, and problems makes other measures more relevant. There are, however two, important additions to make to the methodology that can be explored on the SEIR and SEIS structure. The first extension is to refine the interaction protocols from random mixing to specific mixing patterns. Second, we are interested in gaining more insight into patterns of dynamics by explicitly identifying recurring substructures in the temporal web; i.e., detecting intertemporal motifs. Although the application to simple models is useful for refining and demonstrating the technique, we plan to apply the technique to available temporal network datasets both to evaluate our approach in comparison to others and to provide useful information regarding the behavioral contingencies and dynamical properties.

5.7.1 Underlying Network Structure

The current analysis utilizes homogeneous agents in order to to focus on the ability of tracing the greatest impact agent and time using only the revealed interactions; i.e., a scenario in which no other system information could potentially inform us of the outcome contingencies. In addition to the fully connected base network presented here (which produces uniform random interaction probabilities) we are also interested in exploring the effects of heterogeneous interaction patterns. This can be done using a nonuniform distribution of interaction probabilities or using structured underlying interaction topologies from the literature; for example by using random, small world, scale-free, and ring lattice networks as the potential interaction conduits. This future work will allow us to compare the typical static network properties of agents in the underlying potential interaction structure (and/or the

temporally flattened, observed network structure) to their TKO score and our temporal web measures. Such an analysis will reveal whether an agent's temporally flattened network properties (such as standard degree or betweenness centrality) really operate as reliable proxies for their dynamical properties and system influence.

In principle we could run the static network measures on the social network implied by the temporally flattened graph of the observed connections in these models to test how well standard network measures perform in comparison. Actually we did do that, but with these interaction parameters the temporally flattened graph is nearly fully connected in every run (average edge density is 97%). Standard network analysis is useless here, but exploiting the temporal element fosters further, though different, analysis. Where standard network theory has shined is on more structured social interactions. By (1) restricting interaction along pre-determined pathways, (2) exploring the effects of several different initial conditions on each skeleton, and (3) comparing the actual disease impacts to the impact expected from the standard flat-graph measures we will be able to test the accuracy (and hence usefulness) of these standard measures for propagation dynamics. We may discover, for example, that different network characteristics are important at different stages of the disease spread, and thus different interventions are recommended. This research may also help us identify new measures for flat-graphs that perform better in identifying key agents in social networks.

5.7.2 Dynamical Motif Detection

Network motifs are subgraphs shared across network structures and/or recurring within the same network. Being subgraphs, they are formed from a subset of a network's nodes and their connections, sometimes augmented with particular properties. Structural motifs have been examined in the context of gene regulatory, metabolic, neurological, and many other biological and artificial networks [15–17]. By identifying patterns of importance in the structure of a network, and highlighting all the occurrences of those patterns, motif detection can aid in establishing the building blocks of a larger structure, find key relationships in the structural or functional fabric, fill in missing links, and other descriptive tasks. The idea of using dynamical motifs to capture interesting features of network dynamics appears sporadically in the literature (e.g. [18]), but the approach here is novel to the temporal web construction.

For the temporal web, the motif itself is static and only requires traditional motif detection algorithms to capture and match. The time-directed structure actually simplifies the search algorithm. When the motifs include intertemporal or crosstemporal interaction edges, the motif represents a pattern in the dynamics of the system's state changes and interactions. For example, if we wanted to test

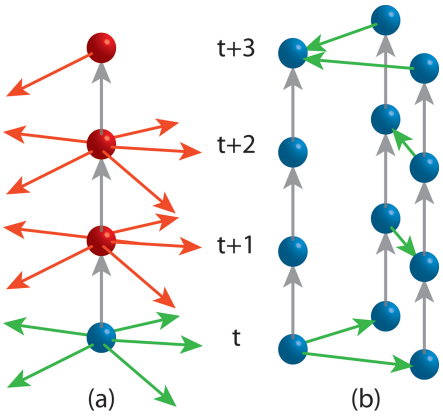


Figure 5.15: Dynamical motifs using intra-temporal relationships representing (a) a reduction in interaction from 5 to 1 occurring two periods after an agent becomes infectious, and (b) a pattern capturing questioning, discussion, and answering (among other behaviors depending on context). Note that the flattened version of (b) is the complete directed non-reflexive graph of three nodes, but this is only one of 4096 possible structures that match the complete flattened graph using a 4-period time span. Motifs are shown here using intra-temporal interactions for ease of presentation, but they can just as easily be defined for cross-temporal interactions.

the hypothesis that people reduce their interaction rate a couple periods after becoming infectious, then this can be represented as a class of motifs in which the degree of nodes is higher before and immediately after becoming infectious than it is two periods later (an example of such a motif is shown in figure 5.15 (a)). Thus hypotheses about individuals' contingent behaviors, or the contingent interactions among several individuals, can be tested for using this technique.

Dynamical motifs may involve multiple elements over time to capture patterns of interaction and changing relationships. If there exists a repeated interaction structure – such as repayment of debt, answering questions, reciprocity, passing along information, etc. – then that will appear as a particular series of connections across time. Collecting the behavioral motifs of a system and comparing them across parameter changes (and even distinct domains) can be used to discern categories of system behavior. By using *motif schemata* (structures including "wildcard" elements similar to those used in genetic algorithms and/or ordinal timing of interaction events) we can eventually generalize recurring themes in system dynamics and establish a catalog of intertemporal interaction patterns. Thus the shared representation as a temporal web facilitates identifying the cross-domain patterns in the dynamics of systems, an important development for the theory of complex systems.

Aside from identifying building blocks and coherent behavior, dynamical mo-

tifs can be used to make predictions over future system behavior. Given a library of known motifs, incomplete patterns can be matched over (for example) the first $\delta t - 1$ periods of the motif. Then one can determine the probability that the final period will complete that motif based on either historical completion frequency or based on all possible next steps (i.e., assuming nothing about likely connection patterns). Such a predictive tool goes beyond exploiting cycles and chains in the network dynamics to basing expectations on a large repertoire of previously observed or theoretically important system behaviors. And because the temporal web can be built from any collection of time-series interaction or relationship data its usefulness extends beyond systems currently modeled as network structures.

5.7.3 Applications to Empirical Data

We are already applying this technique to empirical inter-bank loan data from Russia [19] to detect systemic risk and better understand how it accumulates. We are interested in uncovering specific properties of the interaction system: how do banks mitigate risk, absorb risk, and push risk off to other banks? Under what conditions do bank failures lead to cascades, and are their patterns of behavior that can reliable halt such a cascade. In collaboration with other groups we are also pursuing applications to activity on a brain connectome mapping, Twitter retweet data, online and real life social networks, idea transmission, as well as additional simulation models on a variety of topics. Insights into the behavioral patterns of these systems will come from understanding the dynamics (not the equilibria) and that is why we need improved measures of the dynamical properties.

5.8 Conclusion

We introduced several measures of dynamical properties for identifying the time and agent sensitivity of disease spread in typical SEIR and SEIS models. This simple epidemiological demonstration provides a base from which to explore the dynamical properties of more sophisticated models. The nature of propagation often imbues early infections with the greatest importance for the spreading of a disease in both SEIR and SEIS situations, but not always. Our temporal knockout technique is capable of determining the exact actual contribution of each agent being infected at each time by measuring the change in social morbidity resulting from removing that agent at that time. The out-component and path-related measures perform well overall in identifying the key agents and key times for the disease propagation – and provide quantitative measures of their importance.

For other models the spread of the property of interest will depend much less on the initial agents and more on bottlenecks through the process. This is true of SEIS dynamics after the disease has reached a certain threshold of infection and

becomes sustained in the population through reinfections, and it is true of dissipative systems in general. Our measures may not only perform better in these cases, but also succeed in capturing dynamical properties not yet achievable by any other currently available means. Because the interactions here are chosen randomly, there couldn't be any agent or social property that aligns with their role in the dynamics. However, given the uniformly random interaction structure we still wish to understand the features that determine the sensitivity level of the disease's spread specifically because these are not features of the agents or of the underlying network structure – these are features of the dynamics themselves. Although we have made some encouraging progress on the problem of identifying and measuring these dynamical properties, this is still early work. We expect to make further discoveries in the near future and look forward to collaborating with others on the exploration of temporal webs in its many varieties and applications.

References

[1] Mark Newman. *Networks: an introduction*. Oxford University Press, 2010.

[2] Petter Holme and Jari Saramäkid. *Temporal networks*. Physics Reports, 519(3):97–125, 2012.

[3] Vincenzo Nicosia, John Tang, Cecilia Mascolo, Mirco Musolesi, Giovanni Russo, and Vito Latora. *Graph Metrics for Temporal Networks*. arXiv:1306.0493 [physics.soc-ph], 1306, 2013.

[4] Hartmut H. K. Lentz, Thomas Selhorst, and Igor M. Sokolov. *Unfolding Accessibility Provides a Macroscopic Approach to Temporal Networks*. Physical Review Letters, 110:118701–1 – 5, 2013.

[5] M. E. J. Newman. *Spread of epidemic disease on networks*. Physical ReEview E, 66:016128, 2002.

[6] Hazhir Rahmandad and John Sterman. *Heterogeneity and Network Structure in the Dynamics of Diffusion: Comparing Agent-Based and Differential Equation Models*. Management Science, 54(5):998–1014, May 2008.

[7] René Pfitzner, Ingo Scholtes, Antonios Garas, Claudio J. Tessone, and Frank Schweitzerk. *Betweenness Preference: Quantifying Correlations in the Topological Dynamics of Temporal Networks*. Physical Review Letters, 110:198701 1–5, 2013.

[8] David Thorneycroft, Sarah M. Sherson, and Steven M. Smith. *Using gene knockouts to investigate plant metabolism*. Journal of Experimental Botany, 52(361):1593–1601, 2001.

[9] Stefano Allesina, Antonio Bodini, and Mercedes Pascual. *Functional links and robustness in food webs*. Phil. Trans. R. Soc. B, 364:1701–1709, 2009.

[10] M. Girvan and M. E. J. Newman. *Community structure in social and biological networks*. Proc. Natl. Acad. Sci. USA, 99:7821–7826, 2002.

[11] Ulrik Brandes. *A Faster Algorithm for Betweenness Centrality*. Journal of Mathematical Sociology, 25(2):163–177, 2001.

[12] Marten Scheffer and et al. *Anticipating Critical Transitions*. Science, 338:344–348, 2012.

[13] Alexei Karas and Koen Schoors. *Bank networks, interbank liquidity runs and the identification of banks that are too interconnected to fail*. In Second CInST Banking Workshop, Moscow, 2012.

[14] Dirk Brockmann and Dirk Helbing. *The Hidden Geometry of Complex, Network-Driven Contagion Phenomena*. Science, 342(6164):1337–1342, 2013.

[15] N Kashtan, S. Itzkovitz, R. Milo, and U. Alon. *Topological generalizations of network motifs*. Physical Review E, 70:031909, 2004.

[16] R Milo, S Shen-Orr, S Itzkovitz, N Kashtan, D Chklovskii, and U Alon. *Network motifs: simple building blocks of complex networks*. Science, 298:824?827, 2002.

[17] Olaf Sporns and Rolf Kötter. *Motifs in brain networks*. PLoS Biol, 2(11):e369, 2004.

[18] Valentin Zhigulin. *Dynamical motifs: building blocks of complex dynamics in sparsely connected random networks*. Physical Review Letters, 92(23):238701, 2004.

[19] Alexei Karas and Koen J. L. Schoors. *A Guide to Russian Banks Data*, August 2010.

6

Benchmarking measures of network influence

6.1 Abstract

Identifying key agents for the transmission of diseases (ideas, technology, etc.) across social networks has predominantly relied on measures of centrality on a static base network or a temporally flattened graph of agent interactions. Various measures have been proposed as the best trackers of influence, such as degree centrality, betweenness, and k-shell, depending on the structure of the connectivity. We consider SIR and SIS propagation dynamics on a temporally-extruded network of observed interactions and measure the conditional marginal spread as the change in the magnitude of the infection given the removal of each agent at each time: its temporal knockout (TKO) score. We argue that this TKO score is an effective benchmark measure for evaluating the accuracy of other, often more practical, measures of influence. We find that none of the network measures applied to the induced flat graphs are accurate predictors of network propagation influence on the systems studied; however, temporal networks and the TKO measure provide the requisite targets for the search for effective predictive measures.

6.2 Introduction

In order to effectively prevent the spread of diseases one must identify those individuals with the greatest potential to change propagation outcomes. A similar need exists for efficiently spreading information across a social network. There are two main strategies to identifying the key agents for disease/idea spread: (1) the discovery of "super-spreaders" [1–6] and (2) finding effective immunization/removal targets [7–9]. The difference is not the goal of the analysis; both approaches seek to ascertain the actual or potential *influence* of each node on network propagation by performing an isolated contingency analysis. The first approach is some version of variably seeding an infection and determining how well it spreads in each setup [10, 11]. The second approach is some version of setting nodes as firewalls and measuring changes in how the property/idea/disease spreads with different firewalls [7]. By toggling the status of any one node and examining the differences it generates one can ask, "How much of the propagation is this node responsible for?" Here we propose a measure called "temporal knockout" (TKO) that combines the super-spreader and immunization approaches and also includes the timing of infections to more accurately measure each agent's influence/impact on the propagation.

The dominant technique to assess individual influence is to take a set of agents and a network of potential interactions among them and simulate the propagation of a property using a variation of SI/SIR/SIS dynamics across the network to see how far and how fast it spreads. There are variations in the (generated or empirical) network structure used, the number and placement of initial infections, the disease parameters, and with these there are variations in the identified best measure of influence (see Danon [12] for an extensive review on the possible variations). The most important lesson from these analyses is that different structures make different targets more effective for immunization. For example, connectivity on some network structures is resilient to random node removals but sensitive to targeted removal of nodes with certain properties, such as high degree agents in scale-free networks [13–15]. For other network structures, high degree is not the best measure of importance; betweenness, k-core, and other measures have been proposed as capturing key individuals in certain specific network structures and real-world datasets [16]. Recent work has responded to the inadequacy of traditional centrality measures by developing novel measures such as dynamical influence [17], disease spreading walks [18], accessibility [19], epidemic centrality [6], and expected force [20] among others; although they each share similarities to the common measures or combinations of them.

In order to evaluate network measures' ability to track influence one must have an independent assessment of that influence – the ground truth to be matched. A common way to measure this is to seed the initial infection at each node and

measure the resulting spread, typically as the cumulative cases for SIR. However, an individual's impact on the dynamics of propagation on complex networks is more nuanced than these simple propagation measures indicate. Even when a disease starts at node x, some later-infected node y may be more responsible for the scope of the spread. In actual disease propagation dynamics [21, 22] it is also possible that an agent being infected early *reduces* the eventual scope of the infection by altering the set of individuals that agent comes in contact with while infected.

In light of these possibilities it is clear that one must analyze how the full dynamics unfold in order to correctly assess influence over those dynamics. To incorporate the temporal aspect into our influence analysis we capture the infection propagation in a temporally extruded network structure called a "temporal web" – a variant of temporal networks [23, 24] in which the interactions extend across time creating a single acyclic digraph rather than layered networks [25–27]. This temporal web provides a time-extruded version of cumulative cases that we call "magnitude" combining both the number of infected individuals and the length of their infections [5].

To perform the isolated contingency analysis we propose a measure called "temporal knockout" (TKO) that combines the super-spreader and immunization approaches and also includes the timing of infections to more accurately measure each agent's influence/impact on the propagation. TKO is not an alternative network measure for approximating influence, but rather an all-things-considered empirical measurement of each agent's time-dependent potential to change propagation outcomes for use as a benchmark in evaluating network measures.

First we explain the temporal web construction in more detail, then we describe the process to calculate the disease magnitude and temporal knockout score. Because the temporal knockout score calculation is computationally expensive, it is desirable to have a simpler proxy measure, or set of proxy measures, that accurately reflects agent influence. Toward this end we run a battery of experiments on small world and scale-free networks and evaluate the effectiveness of some standard and newer flat/static network measures to capture influence using the TKO scores as a benchmark measure. The evaluation of network measures presented here is indicative of the need for improved ways to capture propagation influence, but our focus here is the presentation of TKO as a standardized benchmark metric for performing such investigations.

6.3 Approach

Our analysis proceeds through the following steps: (1) create collections of scale-free and small world base networks; (2) build temporal webs encapsulating a fixed set of potential interactions for each one; (3) simulate propagation dynamics across

each temporal web for each agent of each network; (4) calculate the temporal knockout of each node in the temporal web; (5) generate the flattened network and analyze the flat networks using centrality measures; (6) examine the degree to which the flat network measures capture the agents influence as measured by TKO.

6.3.1 Network and Disease Parameters

We simulate the spread of an infectious disease using an agent-based model realizing SIR and SIS dynamics. Our networks have 200 agents connected in either a small world or scale free network with 800 and 784 edges respectively. The small world base networks are undirected connected Watts-Strogatz networks where each agent is connected to $k = 8$ neighbors and the probability of rewiring is set to $p = 0.025$. The scale-free base networks are undirected Barabasi-Albert networks with $m = 4$ as the number of edges to attach from a new node to existing ones. The networks were generated using the implementation of the python package NetworkX [28]. For each combination of network type and infection probability ($p = 0.10$, 0.15, and 0.20), we generate 25 instantiations (150 total). We note that the SIR and SIS versions of a given combination run on the same instantiations, thus using the same link activations at each time step. In each iteration of the model, the probability that a given link is activated is

$$p_{ij} = \frac{k_j^{-1}}{\sum_{K_i} k_n^{-1}}$$

with k_j being the undirected degree of agent j, and the summation in the denominator is over each network neighbor (n) of node i (written K_i). [29] There is one initially-infected agent per run and we perform a run of the model using each agent as the initial agent for each of the 25 instantiations of each network type. Each infectious agent has a probability to infect susceptible network neighbors and we run the full battery of simulations using infection probabilities of 0.10, 0.15, and 0.20. In each period, the probability of infectious agents converting to recovered/susceptible (I→R and I→S for SIR and SIS models respectively) is $1/15$. Each run lasts 200 periods; this is typically sufficient for SIR dynamics to run their course, and is used for SIS models for parsimony of analysis.

6.3.2 Building a Temporal Web

We run our simulations using simultaneous updating so that each agents' state at $t + 1$ depends on their state at t and interactions initiated at t. When represented as an intertemporal network the interaction edges therefore run across time from agents at t to other agents at $t + 1$ in addition to "inheritance edges" from each

agent at t to its $t+1$ self (see Figure 6.1). We call this version of intertemporal networks a "temporal web" because it produces a single acyclic directed graph across time rather than connected layers.

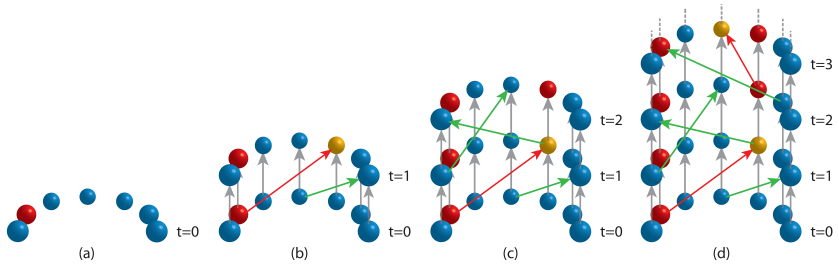


Figure 6.1: A simplified example of building a "temporal web" style intertemporal network from state-change and interaction data for an SIR model. This procedure differs from temporally layered networks in that the interaction edges are cross-temporal to capture simultaneous updating in the generated data, thus creating a single acyclic directed graph across time.

We first build the temporal web "skeleton" that includes all of the state changing and interaction probabilities which may be needed for any particular run. With non-adaptive interaction probabilities, who interacts with whom *and when* all become fixed for those runs. Therefore when we run the simulation using each agent as the initially infected agent, the overall dynamics are kept constant while we monitor the propagation so that the only difference is the initial agent.

6.3.3 Disease Magnitude

The temporal structure facilitates a variety of new measures, which are defined and explored elsewhere [25, 30]. Specifically for epidemiology it becomes natural to switch to a temporally extended refinement of the standard cumulative cases measure. Rather than (or in addition to) reporting the number of agents that are ever infected, the disease *magnitude* is calculated as the number of agent-times (i.e., nodes in the temporal web) that are in the infectious (or exposed) state. It is equivalent to the cumulative sum of the number of infectious agents across iterations [5]. This measure better captures disease morbidity because it accounts for both the number of infections and how long the infections persist – a large number of very short infections could be considered preferable to a few persistent long-term infections. Depending on the application, the node count or a normalized version may be preferable – the number of nodes is the same for all of our experiments described below, so we use the "raw magnitude."

6.3.4 Calculating the Temporal Knockout Scores

Temporal knockout (TKO) measures influence by aggregating two levels of contingency. First we select an agent from the population to be initially infected and run the disease model while capturing each agent's state and interactions at each iteration in a temporal web. The resulting collection of infectious nodes (agent-times) embodies the magnitude of the illness contingent on that agent being the initially infected one. Then the second layer is that for each infected node in the temporal web generated by that run we perform a knockout analysis: remove that node and run *the same infection dynamics* and measure the difference in the disease magnitude. Thus for each node we capture the change in disease magnitude contingent upon that agent being removed at that time, contingent upon that particular initially infected agent.

The initially infected agent at the t_0 iteration will have a marginal infection effect equaling the whole magnitude. Note that removing a noninfectious node at t_0 still prevents it from being infected later, which affects the marginal infection score of that agent at t_0; however, the pre-infection time nodes for an agent will have the same TKO as the first infected time node; thus the calculation can be performed on just the infected subset and backtracked to earlier times. Perhaps counter-intuitively this effect can be negative; i.e., it is possible to remove an agent from the system at a particular time and have the overall disease spread *increase*. This can happen when agents that are infected by the knocked out agent would normally have quickly lead to dead ends, but when instead infected later by other agents they spread the disease to many more others.

We perform this knockout analysis for every node in the temporal web to get the marginal infection score conditional on that initial agent. We repeat this process using each of the agents as the initially infected agent and set each node's TKO score as the average marginal infection score across those runs. Thus we have the conditional marginal infection spread for each agent at each time step for all possible single-agent disease carrier initial conditions. This algorithm therefore captures the *potential* for each agent at each period to influence the spread of the disease.

Because TKO is an overt counting of infected agent-times given the contingent hypothetical-empirical results instead of a summary measure we believe that it stands as a reliable benchmark for the influence of each agent (in networked epidemiological systems). Also note that TKO's hypothetical-empirical approach means that the change in total infection after a knockout of agent A_i at any time t_τ cannot be calculated except through the resimulation of the infection dynamics across the rest of the temporal web. Because of this TKO is thoroughly descriptive of the observed dynamics, but it is not predictive of influence in other runs.

6.3.5 Base and Flattened Graphs

In order to predict which agents are most likely to facilitate diffusion, we wish to compare the TKO identification with measures on flat, non-temporal networks. Specifically we would like to know how well each of various centrality measures does in capturing each agent's network influence as benchmarked by TKO. Two versions of flat graphs are relevant here: (1) the base potential interaction network from which the actual interactions were probabilistically generated and (2) the flattened empirically observed interactions. Our results for the base network and weighted and unweighted flattened networks are nearly identical, so we focus on the base network here and leave the flattened networks for the Supplementary Materials. We have twenty-five distinct base networks for each scenario (although each SIR and SIS pair use the same networks) and for every node in each one we calculate the following centrality and influence measures: k-core, degree, closeness, betweenness, eigenvector, and Katz centralities, accessibility (path length$=2$), and expected force.

6.4 Results

The infection dynamics in our model match other models with similar network structures and disease parameters [12, 29]. We briefly summarize the contagion results in order to provide context for the centrality measures and to facilitate comparisons to other models. For our SIR models the cumulative cases and magnitude measures are nearly perfectly correlated (0.995) because the fixed 1/15 probability of I$\rightarrow$R transitions implies a uniform *expected/average* infection duration time of 15 iterations. For SIS models reinfection can multiply an agent's contribution to magnitude but still only be counted once by the number of cumulative cases, so the correlation is reduced (0.936), but is still high due to the relatively short time horizon for our SIS simulations (200-iterations).

As seen in Table B.1 both network types show high variation in magnitude depending on the initial agent; however, when aggregated across the 25 implementations of each network type they reveal similar magnitude profiles (see supplementary material for details). For ease of reading we present the raw (non-normalized) magnitude scores (i.e., the number of infectious nodes in the temporal web). As you can see in Appendix Figure A.1 there are a large number of runs in which the disease never catches on (what we call "duds") and although these outcomes drag the mean magnitude down and raise the variance, for our purposes there is no benefit in separating out the duds and, for example, testing the remaining infections for matches to known distributions because we do not utilize these summary statistics in any of our TKO analyses.

We also calculate the "epidemic probability" for each agent as a binary vari-

able for whether the run reaches a magnitude greater than 50 when starting at that node. Because, unlike cumulative cases, the magnitude can vary greatly between runs with identical outcomes (e.g., full saturation) due to timing effects the correlation between the magnitude values and epidemic probability is only 0.617. Furthermore, the high variance in magnitude scores (even just among non-dud runs) is large enough to undercut the reasoning for preferring epidemic probability over a fully quantitative measure such as magnitude in this case. [20]

Infection Type	Network Type	Infection Probability	Mean Magnitude	Magnitude StDev	Percent Duds
SIR	Scale Free	0.10	143.352	288.549	0.625
SIR	Scale Free	0.15	584.744	774.628	0.482
SIR	Scale Free	0.20	1296.44	1142.24	0.380
SIR	Small World	0.10	88.9266	131.743	0.584
SIR	Small World	0.15	227.321	324.207	0.457
SIR	Small World	0.20	445.033	559.017	0.352
SIS	Scale Free	0.10	548.746	1155.19	0.593
SIS	Scale Free	0.15	5003.03	5237.44	0.445
SIS	Scale Free	0.20	10800.6	8150.76	0.344
SIS	Small World	0.10	308.734	536.106	0.557
SIS	Small World	0.15	2526.97	2839.26	0.433
SIS	Small World	0.20	7036.79	5623.18	0.333

Table 6.1: Results summary of infection spread for each model variation. Each row aggregates 5000 runs (one run initialized at each of 200 agents for each of the 25 base network implementations). Duds are defined as runs in which the raw magnitude is fewer than 50 agent-times.

6.4.1 TKO vs Magnitude Correlations Results

We first compare the TKO score of each agent to the initial-agent resulting magnitude in order to evaluate whether this standard measure of influence effectively captures a node's ability to spread disease. The TKO algorithm accounts for the idiosyncrasies of the agent interactions across time, but as a result it assigns scores across time as well. In order to compare TKO node scores to initial-agent-spread scores we first need to aggregate them to the individual agents.

For each node we determine two versions of TKO: (1) the proportional change in the number of infectious nodes and (2) the change in the fraction of nodes that become infectious. The proportional change of node i is calculated as the number of agents that are infected when node i has been removed divided by the number of nodes that were originally infected, and then that subtracted from one so that a value of one means that no nodes become infected if this one is removed. Alternatively the delta fraction is the fraction of infected nodes in the original run minus the fraction of nodes that become infected when node i is removed. For both versions negative values occur when more nodes become infected contingent

upon i's removal compared to the original run. An agent that was never infected will have a TKO value of zero for all its temporal nodes. For each of these temporal node-based measures we aggregate them to agents by considering both the maximum value an agent achieves across time and its average TKO score across time.

The Pearson correlations for agent TKO scores and magnitude appear in Table 6.2. In the most correlated scenario (SIR smallworld 0.10 infection rate) the best match is to maximum TKO with a correlation coefficient just under 0.50 (marked with *). Although we initially believed that the Spearman rank correlations would be higher, they are actually very similar and not consistently better or worse (a table of Spearman correlations appears in the Supplementary Materials). For example, the best-case scenario for the Spearman correlation is the same, with a Spearman rho value of 0.517. For both types of correlation the performance drops dramatically as the disease magnitude increases (via higher infection rates), indicating that the large proportion of runs with almost no spread ("duds") are trivially improving the correlations and overstating the ability of agent-initiated magnitude to measure propagation impact.

We also compare agent TKO to epidemic probability and both the Pearson and Spearman correlations of this analysis are very similar overall to the comparison to magnitude, with a maximum Pearson correlation of 0.323 and a maximum Spearman correlation of 0.518 (full tables appear in the Supplementary Materials). Overall they are a weaker match with TKO, but this should be expected because TKO itself is calculated from aggregated contingent marginal magnitudes.

Disease Type	Network Type	InfectionRate	MaxProportion	MaxDeltaFraction	AveProportion	AveDeltaFraction
SIR	scalefree	0.10	0.402831	0.404905	0.288126	0.292081
SIR	scalefree	0.15	0.0674457	0.246309	0.0644679	0.157114
SIR	scalefree	0.20	0.0457449	0.219299	0.0781154	0.158489
SIR	smallworld	0.10	0.494424*	0.471569	0.366466	0.363984
SIR	smallworld	0.15	0.043098	0.264589	0.0770296	0.188738
SIR	smallworld	0.20	0.0297556	0.19209	0.0154106	0.118696
SIS	scalefree	0.10	0.346553	0.375933	0.268197	0.282026
SIS	scalefree	0.15	0.0566606	0.247974	0.0838702	0.153358
SIS	scalefree	0.20	0.0446214	0.233998	0.0585834	0.107849
SIS	smallworld	0.10	0.404204	0.417726	0.353329	0.370916
SIS	smallworld	0.15	0.0240389	0.201376	0.0452133	0.149954
SIS	smallworld	0.20	0.0502245	0.190489	0.0420603	0.10682

Table 6.2: The mean Pearson correlations coefficients across the 25 network instantiations of the disease magnitude given an agent is the initially infected agents and the TKO scores for that agent. The low correlations imply that using the disease spread based on initial infection is a poor measure of influence.

The poor correlations between TKO and both agent-initiated magnitude and epidemic probability have multiple explanations. To understand the relationship better we present a few select plots of the agent TKO scores across time in Figure 6.2. These plots present the change in magnitude resulting from removing each

infectious agent at each time averaged across the 200 runs initialized with each agent being infected. So a value of m means that *on average* (i.e., regardless of which agent is initially infected) removing this agent at this time decreases morbidity by m agent-times.

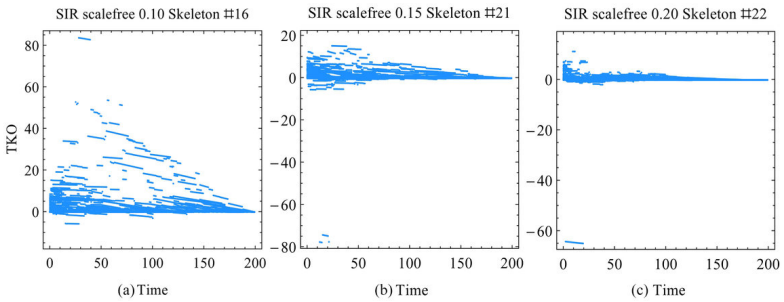


Figure 6.2: Plot of TKO scores across time for SIR dynamics and a scalefree network. These examples shows that the most influential agent-times often do not occur during the initial phases of a disease, but can indicate bottlenecks in the spread of the disease. This also shows the appearance of negative TKO agents, the removal of which actually increases the morbidity of the disease due to timing and network effects.

As we saw, there are many dud runs in which the disease doesn't spread beyond a few initial agents; such cases bring down the average values but they remain comparable across different infection scales here because all our simulations have the same number of runs. A TKO score of twenty might mean 500 saved agent-times in one run and none in the others, or 50 in ten runs, etc. So TKO scores can be small if the disease tends not to spread much because no agent at no time will be a key player in the localized infections. On the other hand, if the disease spreads rapidly from every agent to the whole population then no single agent could be particularly responsible for the scale of the infection across multiple initializations. So TKO will be small in these cases too because there are just too many infection paths for any one agent to be a key player on enough of them to have a high knockout effect. Thus, unlike measuring influence via cumulative cases (or magnitude or epidemic probability) in which every agent may be seen as influential, TKO scores are high only if an agent is influential in the sense of playing a key role in the amount of spread. This difference explains why correlations drop as the infection size increases.

Up to this point we have argued that temporal magnitude is a more accurate measure of disease morbidity than cumulative cases because magnitude accounts for both the length of infection as well as agent reinfections. A network measure's ability to capture an agent's influence on disease is standardly compared to the eventual spread of the disease contingent upon it starting at that agent, but our analysis of correlations with TKO shows that this standard measure of impact itself

fails to capture how much disease spread that agent is responsible for because it lacks sensitivity to the structure of the interactions across time. Specifically, agent-initiated metrics cannot account for the role that some other agent later has in the spread of disease, nor how consistently important a particular agent is for spread regardless of the initial infection. From these results we tentatively conclude that TKO stands as the best measure of an agent's influence on network propagation. We now turn to testing the ability of static network measures to identify a system's high-impact agents.

6.4.2 Predicting Temporal Knockout from the Static Interaction Network

The temporally extruded network structure captures the system dynamics in a way that facilitates contingency analyses, however one must already have the data across time to measure those properties, including TKO. For predictive purposes we would like to know if there is some property of the known interaction structure that can identify key players [8]. Although temporal networks are gaining popularity (see Holme [24] for a review), most network analysis is still performed on flat networks because there are already measures available with known interpretations. The question here is whether any flat graph property can accurately predict the conditional marginal infection as measured by agent-aggregated temporal knockout.

We ran the three comparisons between each of the four aggregated TKO measures and each of eight network centrality measures. Both Pearson and Spearman correlations were calculated. Furthermore, because the standard network centrality measures only purport to capture the highest value agents properly (i.e., rather than a claim to assigning accurate values to all nodes) we also compared the overlap between the ten agents (5%) with the top TKO scores with the ten agents with the top centrality scores [16]. We compared the maximum proportional and maximum delta fraction TKO as well as the average proportional and average delta fraction TKOs with degree, closeness, betweenness, eigenvector, and Katz centrality, as well as accessibility and expected force (k-core values were too undifferentiated on our base networks to be meaningful and are omitted here). The full output of the analysis appears in the Supplementary Materials, but they are qualitatively similar enough that Table 6.3 suffices to understand the general results.

We find that neither the Pearson nor the Spearman correlations are systematically higher, nor is any one of the network measures consistently better than all the others (although eigenvector and Katz centrality are typically worse). Notably,

Disease Type	Network Type	InfectionRate	Degree	Closeness	Betweenness	Eigenvector	Katz	Accessibility	Expected Force
SIR	scalefree	0.10	0.127	0.106	0.099	0.098	0.08	0.081	0.119
SIR	scalefree	0.15	0.125	0.096	0.102	0.098	0.084	0.069	0.114
SIR	scalefree	0.20	0.14	0.119	0.115	0.115	0.101	0.096	0.133
SIR	smallworld	0.10	-0.005	0.031	0.019	-0.008	-0.015	0.039	-0.006
SIR	smallworld	0.15	0.049	0.07	0.082	0.08	0.064	0.091	0.051
SIR	smallworld	0.20	0.067	0.127	0.102	0.026	0.057	0.136	0.062
SIS	scalefree	0.10	0.135	0.089	0.111	0.091	0.068	0.061	0.111
SIS	scalefree	0.15	0.162	0.098	0.127	0.106	0.085	0.054	0.13
SIS	scalefree	0.20	0.233	0.167	0.191	0.168	0.138	0.107	0.201
SIS	smallworld	0.10	0.025	0.054	0.02	-0.013	0.004	0.041	0.022
SIS	smallworld	0.15	0.068	0.102	0.11	0.025	0.056	0.107	0.066
SIS	smallworld	0.20	0.123	0.174	0.241	0.052	0.126	0.232	0.121

Table 6.3: The Pearson correlations between the mean proportional TKO score with each of five base network agent centrality scores. Tables for the other results appear in the Supplementary Materials.

accessibility and expected force, two newer measures specifically designed to measure epidemiological spread, do not fare better than the common centrality measures. Although the correlations are typically positive, the correlation coefficients and Spearman Rhos are almost entirely below 0.20 and there are zero instances across all results of relations above 0.4. Differences between the proportional and delta fraction TKOs are small (as expected), but not negligible; delta fractional correlations tend to be better but not in every case. Similarly, the correlations with mean TKO tend to be slightly higher than maximum TKO, but the differences are small and inconsistent. For the top ten overlap comparison we find that the centrality measures typically find a few of the top ten TKO agents, with the highest average matching score of 0.212.

There are other patterns in the results that may offer clues to where to look for improved network measures. For example, for each disease type, each network type, and each TKO version the correlations of all measures tend to be higher with larger infection rates. Unsurprisingly, degree centrality typically performs better on the scalefree networks than the small world networks. However, any such pattern may be spurious because the correlation values are too low and similar for our sample size to provide adequate power. In summary, the result is that none of the eight measures we consider on the flat interaction network can predict which agents have the greatest influence on spreading a disease.

6.5 Conclusions

In this chapter we have argued that using temporal networks to capture disease spread has the benefits of incorporating the details of the interaction timing which is necessary for judging each agent's level of influence/impact on the spread. The number of infectious agent-time nodes, a measure we call magnitude, is superior to cumulative cases because it captures both the length of infections and agent re-infection. However, adapting the standard measures of influence – eventual spread

contingent upon the starting agent or blocked spread contingent upon removing the agent – to magnitude is insufficient to properly capture an agent's overall level of influence. Although eliminating the initial agent is a sure-fire way to stop the spread, that is not informative for deciding whom to remove before the disease starts. What is needed is the change in the spread of disease contingent upon each agent being removed generalized over all possible initial agents. But the degree of influence is also dependent on *when* the agent is removed because the interaction dynamics of these systems are complex: removing an agent early can increase the eventual spread. We present the temporal knockout measure to capture all these contingencies and provide a general benchmark for propagation influence.

One key insight from this study is that an agent's influence depends on how the dynamics unfold through time, which cannot be accurately predicted by historic interaction data or known communication channels. Nascent measures on temporal network structure (i.e., ones that operate on the full temporal web) can accurately track the TKO property with considerably less computational time, but they still require knowing the complete interaction structure over time [25]. Thus, they work as effective *proxy* measures of TKO on existing temporal webs, but are not viable *predictor* measures of TKO from base graphs. Although we do not have improved static network measures to offer at this stage, we believe that having a proper benchmark for such measures provides the foundation necessary for developing them.

For most realistic health applications, by the time an intervention occurs there are already several infectious individuals, and for this reason there is interest in measures/strategies for scenarios with multiple initially infected agents [12]. The problem is in the combinatorics; e.g., instead of 200 runs per network, with two initial agents it becomes $\binom{200}{2} = 19,900$ runs – for just three initial agents it becomes 1,313,400 runs. Because TKO generalizes marginal conditional spread of every agent-time across all initially infected agents, the TKOs scores can be combined post hoc without needing to rerun the simulations. So, although the TKO algorithm is computationally intense compared to the single initial agent runs, there would be considerable time savings when compared to testing every combination of initially infected agents.

As noted by Kitsak [16], when using cumulative cases to capture the influence of particular agents it makes sense to keep the infection probabilities small enough that the disease typically will not spread to the whole population – otherwise the role of any single individual will be difficult to discern. TKO does not suffer from this limitation because the disease magnitude measure also detects delays in infection even if the whole population does eventually get infected. Again, the timing of the interactions is important, so in addition to facilitating a reduction in morbidity, TKO is useful for developing adaptive intervention strategies.

Recent papers have introduced new measures with claims of increased accu-

racy (at least in certain contexts). However, those accuracy claims are based on how well their own measure matched their own chosen metric on their own chosen network and spread parameters. We propose that TKO, in its exhaustive marginal contingent effect calculation, can act as a benchmark metric against which the performance of proposed measures can be judged – essentially establishing a ground truth for the influence of each agent (at each time) in a network.

We acknowledge that the version of temporal knockout presented here is not the only option for benchmarking epidemiological network studies. One direction of refinement is to develop measures of TKO based on thresholds of infection size changes instead of magnitude – a similar move to using epidemic probability instead of agent-initiated cumulative case. [20, 31] Another direction is to expand the breadth of the simulations to more closely approach an exhaustive analysis of interaction possibilities, perhaps including a notion maintaining high TKO through variations in the infection rate and disease variation into the the measure of influence. [6] We visit these ideas in follow-up research to establish shared benchmarks for evaluating measures of network influence on a variety of standardized generated and empirical networks similar to how Zachary's Karate Club has been used to test community detection methods. Before such benchmark networks can be established, we as a community must agree on what counts as a measure of influence. We propose that temporal knockout may fill that role, and at the very least is a useful step in the right direction.

References

[1] David Kempe, Jon Kleinberg, and Éva Tardos. *Influential nodes in a diffusion model for social networks*. In Automata, languages and programming, pages 1127–1138. Springer, 2005.

[2] Yu Wang, Gao Cong, Guojie Song, and Kunqing Xie. *Community-based greedy algorithm for mining top-k influential nodes in mobile social networks*. In Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining, pages 1039–1048. ACM, 2010.

[3] Masahiro Kimura, Kazumi Saito, Ryohei Nakano, and Hiroshi Motoda. *Extracting influential nodes on a social network for information diffusion*. Data Mining and Knowledge Discovery, 20(1):70–97, 2010.

[4] Duanbing Chen, Linyuan Lü, Ming-Sheng Shang, Yi-Cheng Zhang, and Tao Zhou. *Identifying influential nodes in complex networks*. Physica a: Statistical mechanics and its applications, 391(4):1777–1787, 2012.

[5] Kazumi Saito, Masahiro Kimura, Kouzou Ohara, and Hiroshi Motoda. *Efficient discovery of influential nodes for SIS models in social networks*. Knowledge and information systems, 30(3):613–635, 2012.

[6] Mile Sikic, Alen Lancic, A., Nino Antulov-Fantulin, and Hrvoje Stefancic. *Epidemic centrality – is there an underestimated epidemic impact of network peripheral nodes?* The European Physical Journal B, 86(10):1–13, 2013.

[7] Yiping Chen, Gerald Paul, Shlomo Havlin, Fredrik Liljeros, and H Eugene Stanley. *Finding a better immunization strategy*. Physical review letters, 101(5):058701, 2008.

[8] Yintao Yu, Tanya Y Berger-Wolf, Jared Saia, et al. *Finding spread blockers in dynamic networks*. In Advances in Social Network Mining and Analysis, pages 55–76. Springer, 2010.

[9] Chris J Kuhlman, VS Anil Kumar, Madhav V Marathe, SS Ravi, and Daniel J Rosenkrantz. *Finding critical nodes for inhibiting diffusion of complex contagions in social networks*. In Machine Learning and Knowledge Discovery in Databases, pages 111–127. Springer, 2010.

[10] Mark EJ Newman, Stephanie Forrest, and Justin Balthrop. *Email networks and the spread of computer viruses*. Physical Review E, 66(3):035101, 2002.

[11] A H Dekker. *Network Centrality and Super-Spreaders in Infectious Disease Epidemiology*. In 20th International Congress on Modelling and Simulation (MODSIM2013), 2013.

[12] Leon Danon, Ashley P Ford, Thomas House, Chris P Jewell, Matt J Keeling, Gareth O Roberts, Joshua V Ross, and Matthew C Vernon. *Networks and the epidemiology of infectious disease*. Interdisciplinary perspectives on infectious diseases, 2011, 2011.

[13] Réka Albert, Hawoong Jeong, and Albert-László Barabási. *Error and attack tolerance of complex networks*. Nature, 406:378–382, 2000.

[14] Duncan S. Callaway, M. E. J. Newman, Steven H. Strogatz, and Duncan J. Watts. *Network Robustness and Fragility: Percolation on Random Graphs*. Physical Review Letters, 85, 25:5468–5471, 2000.

[15] R. Pastor-Satorras and A. Vespignani. *Immunization of complex networks*. Physical Review E, 65:036104, 2002.

[16] Maksim Kitsak, Lazaros K. Gallos, Shlomo Havlin, Fredrik Liljeros, Lev Muchnik, H. Eugene Stanley, and Hernán A. Makse. *Identification of influential spreaders in complex networks*. Nature Physics, 6:888–893, 2010.

[17] Konstantin Klemm, M Serrano, Víctor M Eguíluz, and Maxi San Miguel. *A measure of individual role in collective dynamics*. arXiv preprint arXiv:1002.4042, 2010.

[18] Frank Bauer and Joseph T Lizier. *Identifying influential spreaders and efficiently estimating infection numbers in epidemic models: A walk counting approach*. EPL (Europhysics Letters), 99(6):68007, 2012.

[19] Matheus P Viana, João LB Batista, and Luciano da F Costa. *Effective number of accessed nodes in complex networks*. Physical Review E, 85(3):036105, 2012.

[20] Glenn Lawyer. *Understanding the influence of all nodes in a network*. Scientific reports, 5, 2015.

[21] Lars Hufnagel, Dirk Brockmann, and Theo Geisel. *Forecast and control of epidemics in a globalized world*. Proceedings of the National Academy of Sciences of the United States of America, 101(42):15124–15129, 2004.

[22] Dirk Brockmann and Dirk Helbing. *The Hidden Geometry of Complex, Network-Driven Contagion Phenomena*. Science, 342(6164):1337–1342, 2013.

[23] Petter Holme and Jari Saramäkid. *Temporal networks*. Physics Reports, 519(3):97–125, 2012.

[24] Petter Holme. *Modern temporal network theory: a colloquium*. The European Physical Journal B, 88(9):1–30, 2015.

[25] Aaron Bramson and Benjamin Vandermarliere. *Dynamical properties of interaction data*. Journal of Complex Networks, 4(1):87–114, 2016.

[26] Othon Michail. *An introduction to temporal graphs: An algorithmic perspective*. In Algorithms, Probability, Networks, and Games, pages 308–343. Springer, 2015.

[27] Leo Speidel, Taro Takaguchi, and Naoki Masuda. *Community detection in directed acyclic graphs*. The European Physical Journal B, 88(8):1–10, 2015.

[28] Aric A. Hagberg, Daniel A. Schult, and Pieter J. Swart. *Exploring network structure, dynamics, and function using NetworkX*. In Proceedings of the 7th Python in Science Conference (SciPy2008), pages 11–15, Pasadena, CA USA, August 2008.

[29] Hazhir Rahmandad and John Sterman. *Heterogeneity and Network Structure in the Dynamics of Diffusion: Comparing Agent-Based and Differential Equation Models*. Management Science, 54(5):998–1014, May 2008.

[30] René Pfitzner, Ingo Scholtes, Antonios Garas, Claudio J. Tessone, and Frank Schweitzerk. *Betweenness Preference: Quantifying Correlations in the Topological Dynamics of Temporal Networks*. Physical Review Letters, 110:198701 1–5, 2013.

[31] Alen Lancic, Nino Antulov-Fantulin, Mile Sikic, and Hrvoje Stefancic. *Phase diagram of epidemic spreading – unimodal vs. bimodal probability distributions*. Physica A: Statistical Mechanics and its Applications, 390(1):65–76, 2011.

7

Detection and localization of change points in temporal networks with the aid of stochastic block models

7.1 Abstract

A framework based on generalized hierarchical random graphs (GHRGs) for the detection of change points in the structure of temporal networks has recently been developed by Peel and Clauset [1]. We build on this methodology and extend it to also include the versatile stochastic block models (SBMs) as a parametric family for reconstructing the empirical networks. We use five different techniques for change point detection on prototypical temporal networks, including empirical and synthetic ones. We find that none of the considered methods can consistently outperform the others when it comes to detecting and locating the expected change points in empirical temporal networks. With respect to the precision and the recall of the results of the change points, we find that the method based on a degree-corrected SBM has better recall properties than other dedicated methods, especially for sparse networks and smaller sliding time window widths.

7.2 Introduction

Networks are currently widely used to map and study interacting systems of animate and inanimate objects [2–4]. Often, the methodologies and measures developed within the context of network theories allow one to identify the central players [5–7] and to find structures in the nodal interactions of the network [8]. Thereby one often identifies groups of nodes - communities - which interact more within a group than across groups [9–11]. Other frequently obtained topologies of social networks include the core-periphery structure [12, 13] with a small group of highly interconnected core nodes and a large group of peripheral nodes that do mostly interact with core nodes.

As the dynamical origins of the interactions evolve over time, the topology of the network can change [14, 15]. For example, a social network of high-school students changes between "normal classes" mode and "summer break" mode, not to speak about what happens to the network after graduation [16]. There are many time evolving networks, however, for which the identification of the changes in the topological structure of the network is not that obvious. Recently, Peel and Clauset [1] proposed a framework to locate the structural breaks in the large-scale structure of time-evolving networks. The proposed change point detection methodology of [1] develops in four steps:

(1) Select the generalized hierarchical random graph (GHRG) parametric family of probability distributions appropriate for reconstruction of the empirical network data.

(2) Select an appropriate width w of a sliding time window.

(3) For each time window, use the proposed parametric family of probability distributions to infer two versions for the model: one corresponding with a change of parameters at a particular instance of time within the window, and an alternate one corresponding with the null hypothesis of no change point over the entire time window.

(4) Conduct a statistical hypothesis test to determine whether the "change" or "no-change" mode provides the better fit to the empirical network data.

In this paper we build on this methodology, but introduce also stochastic block models (SBM) as a parametric family for reconstructing the empirical network in step (1) of the above-mentioned procedure. The SBMs have the advantage of being very flexible. Indeed, they can capture for example both assortative and disassortative behaviour, and core-periphery networks [17–19]. An alternate method for change point detection with an adaptive time window based on Markov chain Monte Carlo and SBMs has recently been outlined in [16].

In what follows, we first introduce the concept of SBMs to capture a given empirical network. Next, we detail a new method to fit a model to a given empirical network and to find the change points in a sliding time window of size w. In section 7.5 we apply our proposed methodology to a number of prototypical temporal networks. We introduce several strategies to detect change points and compare the quality of their results. First, we conduct a study with synthetic temporal networks. Next, we apply the change-point detection methods to three empirical temporal social networks: the Enron e-mail network, the MIT proximity network, and the international trade network after 1870. For these three networks the empirical change points are documented and we compare those with the numerical predictions.

7.3 Fitting stochastic block models to a network

In its simplest form, an SBM distributes the N nodes of a network into K groups. With n_r we denote the prior probability that a node is classified in group r. Obviously, one has that $\sum_r n_r = 1$. Let $\mathcal{Q}_{rs}$ be the probability that a link exists between a node u in block r and a node v in block s. The parameters $\mathcal{Q}_{rs}$ form a $K \times K$ matrix ($1 \leq K \leq N$). We call $g_u = r$ ($g_v = s$) the block assigned to node u (v). With these conventions, the probability of having a link between nodes u and v is Bernoulli distributed with parameter $\mathcal{Q}_{g_u g_v}$. One can determine the likelihood of a given network (as fully determined by its adjacency matrix A) with a given node partitioning $\{g_u\}$ given the SBM model parameters $\{n_r\}$ and $\{\mathcal{Q}_{rs}\}$. This can be expressed either in terms of a product over all nodes, or in terms of a product over all blocks.

$$\begin{aligned} P(A, \{g_u\}|\{n_r\}, \{\mathcal{Q}_{rs}\}) &= \prod_u n_{g_u} \prod_{u<v} \mathcal{Q}_{g_u g_v}^{A_{uv}} \left(1 - \mathcal{Q}_{g_u g_v}\right)^{1-A_{uv}} \\ &= \prod_r n_r^{N_r} \prod_{r \leq s} \mathcal{Q}_{rs}^{m_{rs}} \left(1 - \mathcal{Q}_{rs}\right)^{N_{rs} - m_{rs}}. \end{aligned} \tag{7.1}$$

Here, m_{rs} is the number of actual links between nodes in block r and nodes in block s. Further, N_{rs} is the total number of possible links between the nodes in block r and the nodes in block s. For multigraphs, where A_{uv} can be larger than one, the distributions in the right-hand-sides of (7.1) can be replaced by Poisson distributions. One finds for the multigraph versions of the likelihood of (7.1)

$$\begin{aligned} P^{(\text{Poisson})}(A, \{g_u\}|\{n_r\}, \{\mathcal{Q}_{rs}\}) &= \prod_u n_{g_u} \prod_{u<v} \frac{\mathcal{Q}_{g_u g_v}^{A_{uv}} e^{-\mathcal{Q}_{g_u g_v}}}{A_{uv}!} \\ &= \prod_r n_r^{N_r} \prod_{r \leq s} \mathcal{Q}_{rs}^{m_{rs}} e^{-N_{rs}\mathcal{Q}_{rs}} \prod_{u<v} \frac{1}{A_{uv}!}. \end{aligned} \tag{7.2}$$

These expressions for the probability distributions make the SBM a powerful and versatile tool for the analysis of complex networks.

With the eye on community detection in networks, one often uses the degree-corrected (DC) version of SBM [19]. Thereby, one introduces for all nodes u an extra parameter θ_u proportional to the ratio of u's degree to the sum of all degrees in block g_u. By doing so, the link probability $\mathcal{Q}_{g_u g_v}$ can be replaced by $\mathcal{Q}_{g_u g_v}\theta_u\theta_v$ as the probability for a link between nodes u and v. This replacement diminishes the dependence of $\mathcal{Q}_{g_u g_v}$ on the magnitude of the degrees of nodes u and v. As a consequence, the likelihood that a node with low degree and a node with high degree belong to the same group increases, provided that their θ is low and high, respectively. The sketched degree correction makes sure that a separation into modules is more likely than a separation into groups with similar degrees as often happens with the regular SBM version. We refer to [19] for more details concerning the degree correction.

We now detail our proposed method to fit a parametric distribution to a given empirical network. As in [20, 21] we use belief propagation to fit an SBM to a given network. Thereby, each node u sends a "message" $\psi_r^{u\to v}$ to every other node v in the network. The $\psi_r^{u\to v}$ indicates the probability that node u would belong to block r, in the absence of node v. These conditional probabilities can be iteratively updated with the aid of the expression

$$\psi_r^{u\to v} = \frac{1}{Z^{u\to v}} n_r \prod_{w\neq u,v} \left(\sum_s P(A_{wu}|\mathcal{Q}_{sr})\psi_s^{w\to u} \right), \tag{7.3}$$

with the normalization coefficient,

$$Z^{u\to v} = \sum_{rs} P(A_{uv}|\mathcal{Q}_{rs})\psi_r^{u\to v}\psi_s^{v\to u}. \tag{7.4}$$

The marginal probability ψ_r^u that node u belongs to block r can then be obtained from the following expression

$$\psi_r^u = \frac{1}{Z^u} n_r \prod_{w\neq u} \left(\sum_s P(A_{wu}|\mathcal{Q}_{sr})\psi_s^{w\to u} \right), \tag{7.5}$$

with the normalization coefficient

$$Z^u = \sum_r n_r \prod_{w\neq u} \left(\sum_s P(A_{wu}|\mathcal{Q}_{sr})\psi_s^{w\to u} \right). \tag{7.6}$$

In order to make the algorithm scalable, it is worth remarking that up to $\mathcal{O}(\frac{1}{N})$ terms, the "messages" between two unconnected nodes (u, v) (with $A_{uv} = 0$) can be approximated by the marginal probability (see [20] for details)

$$\psi_r^{u\to v} \approx \psi_r^u. \tag{7.7}$$

With this approximation, for each node u one stores and updates the ψ_r^u and the $\psi_r^{u\to v}$ for u's neighbours $\{v|v \neq u, A_{uv} > 0\}$. This reduces the number of "messages" to be updated to $N + M$, with M the total number of links in the network. Without the approximation (7.7), N^2 probabilities $\psi_r^{u\to v}$ need to updated and stored.

The "messages" of (7.3) and (7.5) allow one to put forward estimates of the SBM parameters

$$n_r = \left\langle \frac{N_r}{N} \right\rangle = \frac{\sum_u \psi_r^u}{N}, \tag{7.8}$$

$$\begin{aligned}\mathcal{Q}_{rs} &= \left\langle \frac{m_{rs}}{N_{rs}} \right\rangle \\ &= \begin{cases} \frac{1}{N^2\left(\sum_{u'} \psi_r^{u'}\right)\left(\sum_{v'} \psi_s^{v'}\right)} \sum_{u\neq v} \frac{A_{uv} P(A_{uv}|\mathcal{Q}_{rs}) \psi_r^{u\to v} \psi_s^{v\to u}}{Z^{uv}} & (r \neq s) \\ \frac{1}{N^2\left(\sum_{u'} \psi_r^{u'}\right)\left(\left(\sum_{v'} \psi_s^{v'}\right) - 1/N\right)} \sum_{u\neq v} \frac{A_{uv} P(A_{uv}|\mathcal{Q}_{rs}) \psi_r^{u\to v} \psi_s^{v\to u}}{Z^{uv}} & (r = s). \end{cases}\end{aligned} \tag{7.9}$$

Using (7.3) one can update the "messages" $\{\psi_r^{u\to v}\}$ given the current estimates of the SBM parameters $\{n_r\}$ and $\{\mathcal{Q}_{rs}\}$. The expressions (7.8) and (7.9), on the other hand, provide a way to estimate the SBM parameters, given the "messages". Fitting the SBM to an empirical network can then be done as follows:

(1) Initialise $\{\psi_r^{u\to v}\}$ for each node u, and the parameters $\{n_r\}$ and $\{\mathcal{Q}_{rs}\}$ randomly.

(2) Update the SBM parameters using (7.8) and (7.9).

(3) Iteratively update the "messages" $\{\psi_r^{u\to v}\}$ and $\{\psi_r^u\}$, using (7.3) and (7.5) respectively, until they converge.

(4) Repeat steps (2) and (3) until both the parameters $(\{n_r\}, \{\mathcal{Q}_{rs}\})$ and the "messages" $(\{\psi_r^{u\to v}\}, \{\psi_r^u\})$ have converged.

This is a variant of the Expectation-Maximisation algorithm that finds the optimal parameter values using point estimates for a given initialisation. Because this approach can cause convergence to a local minimum, it is safer to execute this algorithm multiple times with different random initialisations, and accept the solution with the highest likelihood.

By using (7.8) and (7.9) we obtain estimates of the network's parameters of which we deem that they offer some advantages over an approach that assigns the nodes to blocks deterministically. This is because a node u that has a high probability to reside in block r ($\psi_r^u \simeq 1$), retains a small probability of residing in block $s \neq r$ ($\psi_s^u > 0$). Accordingly, it contributes to the estimate of $\mathcal{Q}_{ss}$ through (7.9). This avoids the following problem that occurs with the deterministic assignment of

the nodes to blocks. Suppose that a block s has a deterministically assigned set of nodes. In situations whereby those nodes have no links in the underlying network, $\mathcal{Q}_{ss}$ is estimated as zero. By the same token, using (7.1) or (7.2) the likelihood of a network with one link in block s is also zero. In the approach adopted in this work, the estimate of $\mathcal{Q}_{ss}$ differs from zero which implies that the likelihood of a link within block s differs from zero. Indeed, this is guaranteed through the use of (7.8) and (7.9), and the fact that $\psi_s^u > 0$ for all or nearly all nodes u. An alternate way of circumventing the sketched problem is to introduce Bayesian priors for the ψ_s^u, as was done in [1].

We now discuss the method used to determine the number of blocks K. To this end, we repeat the above fitting procedure for various choices of K, and select the one with the minimum description length (DL). We use the definition of the DL proposed in [22]. It consists of the sum of an entropy term $\mathcal{S}$ accounting for the amount of information in the network that is described by the model, and of a model information term $\mathcal{L}$ that quantifies the information needed to describe the model. After a deterministic assignment of the nodes to blocks using $g_u = \arg\max_r \psi_r^u$, the DL Σ can be written as:

$$\begin{aligned}\Sigma = &\sum_r \ln \left(\!\!\binom{N_r}{2}\!\!\right) + \sum_{r<s} \ln \left(\!\!\binom{N_r N_s}{m_{rs}}\!\!\right) \\ &+ \ln \left(\!\!\binom{\left(\!\!\binom{K}{2}\!\!\right)}{M}\!\!\right) + \ln \left(\!\!\binom{K}{M}\!\!\right) + \ln N! - \sum_r N_r \ , \end{aligned} \tag{7.10}$$

where $\left(\!\binom{N}{m}\!\right) = \binom{N+m-1}{m}$ is a combination with repetitions. For directed networks, the first line of (7.10) becomes $\sum_{rs} \ln \left(\!\binom{N_r N_s}{m_{rs}}\!\right)$. For the degree-corrected model, and for more information on the MDL for SBMs, we refer to [18] and [22]. In particular, Appendix A of [22] points out that the use of the MDL is equivalent to a Bayesian model selection of the parameter K.

7.4 Method for detection and localization of change points

In this work, we define a temporal network as a time series of consecutive snapshots of a network. Using the methodology of the previous section to fit an SBM to a given network, we can now proceed to develop a technique appropriate for the detection of change points in a temporal network. The methodology rests on the idea to use an overlapping sliding time window with width w and to statistically determine for each time window whether it contains a change point or not. With this procedure, one can detect change points without taking the full time series of networks into consideration.

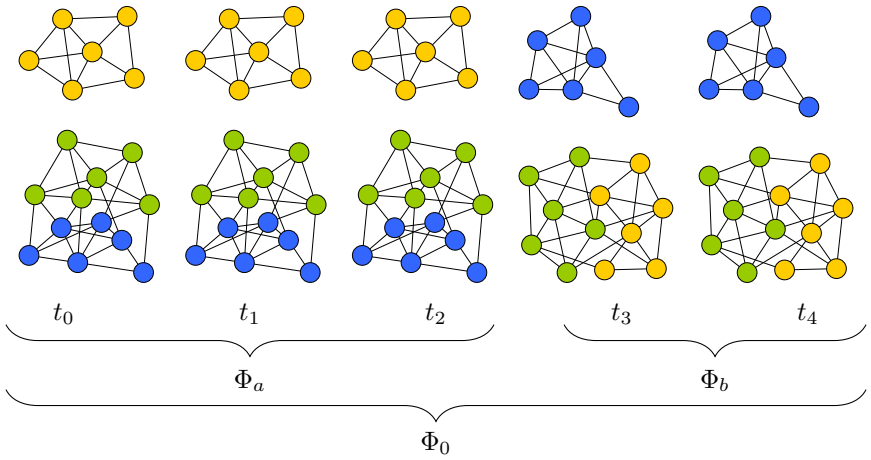


Figure 7.1: Window of five consecutive snapshots $(t_0, t_1, t_2, t_3, t_4)$ of a temporal network containing a change point between t_2 and t_3. Before the change point there are two distinct communities. After the change point the green nodes change sides and now make up a new community with the yellow nodes. Model Φ_0 represents the null hypothesis that there is no change point in the considered time window. A change point is detected when the combination of the two models Φ_a (fit to (t_0, t_1, t_2)) and Φ_b (fit to (t_3, t_4)) are statistically identified as a better fit to the empirical network data at five time instances.

For each time window, we label the graphs by means of their time coordinate $(t_0, t_1, \ldots, t_{w-1})$ (see figure 7.1). We can test the hypothesis that a change point occurs in a particular window by considering all $w - 1$ times $(t_1, \ldots, t_{w-1})$ as possible change points. Of those the most likely one is selected. As a basis of reference, we start from the null hypothesis of no occurrence of a change point in the considered time window. This hypothesis assumes no change point in the window of networks, and can therefore be based on an average model for all the networks in the window. In order to construct such an average model in a given time window, we proceed as follows:

(1) In any given time window, add all the links between every pair (u, v) of nodes and construct $A_{uv}^{[0,w-1]} = \sum_{t=t_0}^{t_{w-1}} A_{uv}^t$. This then forms a multigraph or a weighted network with discrete weights $0 \leq A_{uv}^{[0,w-1]} \leq w$.

(2) Using (7.2) a Poisson-distributed SBM is fitted to the obtained multigraph $A_{uv}^{[0,w-1]}$ using the belief propagation technique detailed in the previous section. Thereafter, the corresponding parameters $\{\mathcal{Q}_{rs}\}$ are divided by the window length w. This ensures that the expected number of links between two nodes is the average number for all network realisations in the window, rather than the sum.

This model then forms the null model Φ_0 in a conventional likelihood-ratio test. The alternative hypothesis states that a change point occurs just before the network realisation at time instant t_n, with $t_0 < t_n < t_w$. For the alternative hypothesis, two other models can be constructed by re-estimating the SBM-parameters for the networks before t_n (resulting in Φ_a), and for the networks from t_n on (resulting in Φ_b) (Figure 7.1). There are $w - 1$ such hypotheses, each of which results in a log-likelihood ratio

$$\Lambda_{t_n} = \sum_{t=t_0}^{t_n-1} \ln P(A^t|\Phi_a) + \sum_{t=t_n}^{t_0+w-1} \ln P(A^t|\Phi_b) - \sum_{t=t_0}^{t_0+w-1} \ln P(A^t|\Phi_0). \quad (7.11)$$

In order to determine the potential change point t_n we select the maximum of these log-likelihood ratios,

$$g = \max_{t_n} \Lambda_{t_n} \ . \quad (7.12)$$

What remains to be done is to determine whether the potential change point t_n is significant. This selection can be done by choosing a threshold value for g. The traditional method to model the distribution of the log-likelihood ratios, using Wilks' theorem, is with a χ^2-distribution. It has been shown [21], however, that this asymptotic approximation does not apply to a SBM. Therefore, as in [1], we make use of bootstrapping. Bootstrapping is a way to model the distribution of the log-likelihood ratio for windows that fall under the null model, called the null distribution. This is achieved by generating a large number of networks from the null model, and calculating the log-likelihood ratio g' using (7.12) for every w of these networks. As for these networks no change point should be detected, these $\{g'\}$ can be assumed to be samples from the distribution of the null model. We can then use the distribution of these $\{g'\}$ as an approximation of the real null distribution. A decision for the detection of a change point can then be made by selecting a confidence level and corresponding significance level, e.g. $1 - \alpha = 95\%$. We calculate the p-value of the log-likelihood ratio g as

$$p = \frac{|\{g'\} > g|}{|\{g'\}|}. \quad (7.13)$$

The p-value determines the significance of the log-likelihood ratio, and the change point is only accepted if the condition $p < \alpha$ is met.

7.5 Results

In this section we present the results of our numerical studies of change-point detection. We use both synthetic (section 7.5.1) and empirical (section 7.5.2) temporal networks. For all those temporal networks we use in total five methodologies to detect and locate the change points. First, the degree-corrected and the regular

SBM techniques introduced in this work (DC-SBM, SBM) and the GHRG method introduced in [1]. We confront the results of those three involving methodologies with those of two rather straightforward local methods based on the mean degree and mean geodesic of the network. For these local methods, we calculate the specified scalars for each network in a given time window and for the network at the time instance just after the considered time window. The value for this last network is then compared to the mean value for the networks in the window, by means of a two-tailed Student's t-test. Thereby we adopt the same significance level α as used for the other methods ($1 - \alpha = 95\%$).

7.5.1 Analysis with synthetic temporal networks

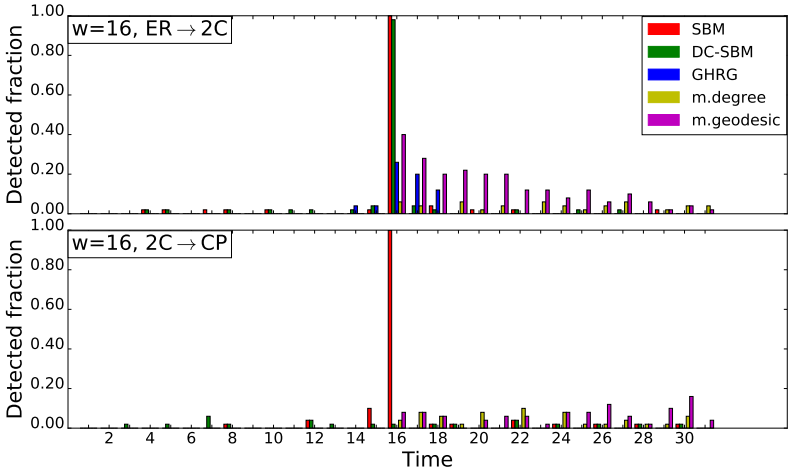


Figure 7.2: The efficiency of detecting a change point in two synthetic temporal networks with the SBM, DC-SBM, GHRG, mean-degree, and mean-geodesic methods. The true location of the change point is $t = 16$. Upper panel: $t = 16$ marks the change from an Erdős-Rényi (ER) network to a network with two communities (2C). Lower panel: $t = 16$ marks the change from a network with two communities to a network with a core-periphery (CP) structure. At all time instances, the height of the bar indicates the fraction of the 50 simulations that detect a change point. A sliding time window of size $w = 16$ was used.

In this subsection we compare the performance of the proposed techniques at the retrieval of planted change points in synthetic temporal networks. We apply the methodology outlined in Sections 7.3 and 7.4 to the synthetic transition from an Erdős-Rényi (ER) network into a network with two communities (2C), and from a network with two communities into a network with a core-periphery (CP) structure.

We report results of four rounds of studies each covering 50 simulations of

32 time instances. Thereby, the change point is planted at $t = 16$. The temporal synthetic networks of the "ER", "2C" and "CP" type are generated from their defining SBMs, with a fixed number of nodes in each block. More specifically, the results reported are generated from:

$$\text{ER} \to \text{2C:} \begin{pmatrix} 0.1 & 0.1 \\ 0.1 & 0.1 \end{pmatrix} \to \begin{pmatrix} 0.15 & 0.05 \\ 0.05 & 0.15 \end{pmatrix}, \qquad \overline{N} = \begin{pmatrix} 22 \\ 28 \end{pmatrix}, \quad (7.14)$$

$$\text{2C} \to \text{CP:} \begin{pmatrix} 0.2 & 0.01 \\ 0.01 & 0.2 \end{pmatrix} \to \begin{pmatrix} 0.3 & 0.09 \\ 0.09 & 0.01 \end{pmatrix}, \quad \overline{N} = \begin{pmatrix} 20 \\ 30 \end{pmatrix}, \quad (7.15)$$

$$\text{CP} \to \text{2C:} \begin{pmatrix} 0.3 & 0.09 \\ 0.09 & 0.01 \end{pmatrix} \to \begin{pmatrix} 0.2 & 0.01 \\ 0.01 & 0.2 \end{pmatrix}, \quad \overline{N} = \begin{pmatrix} 20 \\ 30 \end{pmatrix}. \quad (7.16)$$

For each simulation of a given set-up, 16 networks are independently generated from the first SBM, followed by 16 independent networks from the second SBM. This creates a time series of networks with larger variations than those typically found in the empirical temporal networks that will constitute the study of Sec. 7.5.2. We stress that the GHRG model would be an equally good choice to generate the synthetic temporal networks.

Figure 7.2 summarizes the results of the detection efficiencies for the "ER→2C" and "2C→CP" transitions, using a sliding window of size $w = 16$, and a significance level of $1-\alpha = 95\,\%$. We observe that the regular SBM method (and for the formation of two communities also the DC-SBM method) has a very high detection rate at the change point. The GHRG and the local methods have a significantly lower detection rate.

Figure 7.3 shows the change-point detection efficiencies for two transitions related to those of Fig. 7.2. The first is the ER→2C transition with a window size $w = 4$. Comparison to the upper panel in Fig. 7.2 illustrates that a shorter window size causes the SBM method to predict more false predictions for local change points. We stress that those can be partially attributed to the adopted algorithm that generates the networks independently. Figure 7.3 also shows the detection efficiency results for the CP→2C transition with $w = 16$. This is the time reversed process of the one shown in the lower panel of Fig. 7.2. The DC-SBM method shows a noticeable increase in detections of a change point. This is in line with the expectations, as the DC-SBM is more adept at discovering community structure than the regular SBM. This indicates that the DC-SBM method is better at discovering the formation of a community structure than it is at discovering its dissolution.

In the studies summarized in Figs. 7.2 and 7.3 the GHRG method seems to under-perform. The underlying reasons can be understood by inspecting Fig. 7.4 showing for one specific studied transition the mean of one minus the p-value of the likelihood ratio statistic, which can be interpreted as the probability of occur-

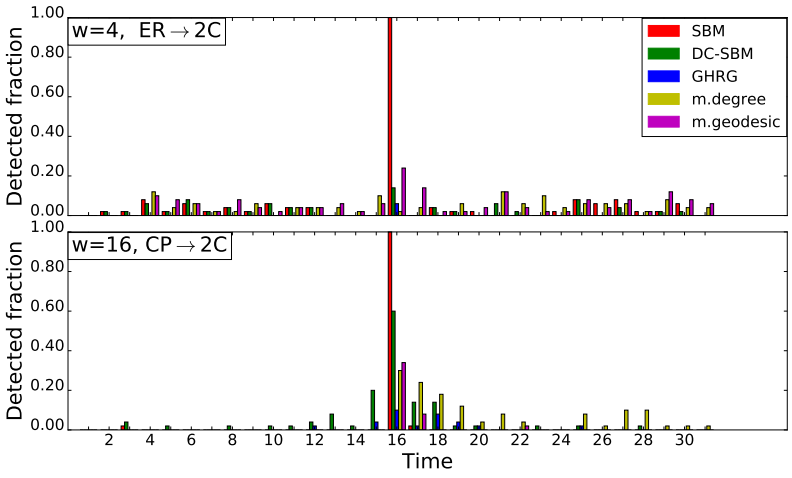


Figure 7.3: As in Fig. 7.2 but for a different value of w (upper panel) and for the time reversed process (bottom panel).

rence of a change point. We see that the GHRG, like the other methods, produces a peak in this probability, centred around the real change point. The mean, however, doesn't rise above the 95 % that was put forward as the detection threshold. This indicates that at lower values of this threshold, the GHRG method would be equally efficient at predicting the $t = 16$ peak. We stress that similar observations are made for all the transitions considered.

7.5.2 Analysis with empirical temporal networks

We now apply the methodology outlined in the Sections 7.3 and 7.4 to three empirical temporal networks: the Enron e-mail network, the MIT proximity network and the international trade network. The first two datasets were also used in the change point analysis of [1]. First, we briefly describe the three datasets that underlie the temporal networks used in our analysis.

Enron is a U.S. energy company that filed for bankruptcy back in 2001 due to accounting scandals. As a result of an official inquiry, a dataset of e-mails exchanged between members of the Enron staff was made public[1]. With those data, one can construct a temporal network with Enron's staff members as nodes, and links which reflect the e-mail exchanges in a particular working week. In this way, one creates a sparse network with an average of 0.43 links per node.

The MIT reality mining project is an experiment conducted by the Media Laboratory at the Massachusetts Institute of Technology (MIT) during the 2004-2005

[1] Available at `www.cs.cmu.edu/~enron`.

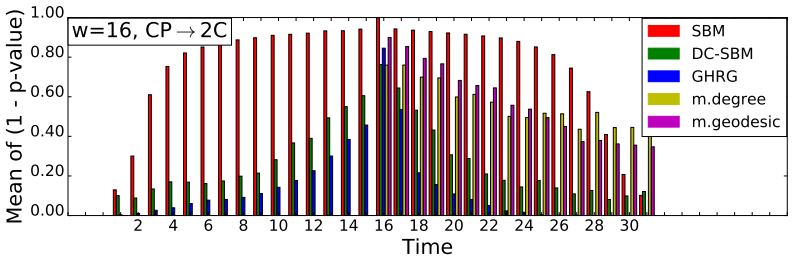


Figure 7.4: The estimated probability of detecting a change point in a synthetic temporal network with the SBM, DC-SBM, GHRG, mean-degree, and mean-geodesic methods. The true location of the change point is $t = 16$. It marks the change from a network with a core-periphery (CP) structure to a network with two communities. At all time instances, the height of the bar indicates one minus the p-value of the likelihood ratio statistic, averaged over all time windows containing the candidate change point for the SBM, DC-SBM and GHRG methods, and over the 50 simulations. A sliding time window of size $w = 16$ was used.

academic year [23]. In this experiment, ninety-four subjects, both MIT students and staff, were monitored by means of their smartphone. Thereby, the Bluetooth data give a measure of the proximity between two subjects[2]. This proximity can be interpreted as a link between two subjects. As the time of proximity is also recorded, one can produce a weekly empirical temporal network by grouping the links per week. In this way, a dense network with an average of 9.07 links per node is obtained.

The study of international trade before the 1950s is hampered by the limitations imposed by the scarcity of data. Thanks to a technique developed in [24], a reliable coverage of the data on international trade between 1880 and 2011 could be accomplished. Note that during the world wars data collection on trade was almost halted. Hence, we exclude these periods from the sample. We construct a temporal international trade network with countries as nodes and establishing links whenever the countries have a significant level of trade integration in a specific year. We treat the international trade data as undirected in order to make a change point analysis with the GHRG method possible.

For the Enron e-mail and MIT proximity networks we consider all nodes (including those with no links) in the time windows. For the international trade network, however, we retain the nodes with at least one link throughout the window. In this way a more dense network is obtained, creating improved conditions for change point detection.

For each of the three considered temporal networks, there are a number of known dates corresponding with events that are likely to have impacted the net-

[2]Available at `http://realitycommons.media.mit.edu/realitymining.html`.

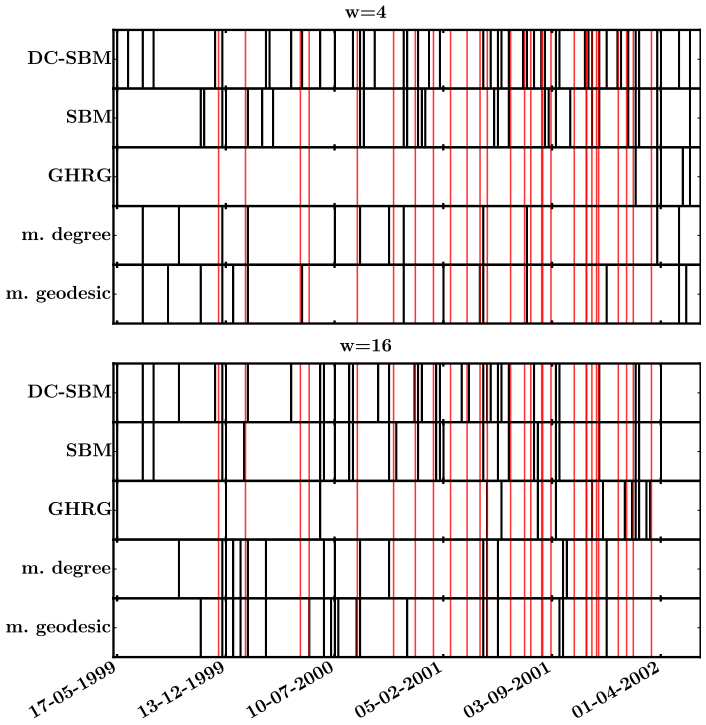


Figure 7.5: The detected change points in the Enron e-mail network for $w = 4$ weeks (upper panel) and $w = 16$ weeks (lower panel). Use has been made of the SBM, DC-SBM, GHRG, mean-degree and mean-geodesic methods. The red vertical lines correspond with the time instances of documented events in the Enron company.

work's structure. We treat those dates as if they were the "empirical" change points, realizing that they merely mark dates with an enhanced likelihood for changes in the network to occur. The major purpose of the introduction of "empirical" change points is to develop a quantitative measure to compare the figure of merit of the different change point detection methodologies. In order to quantify the quality of the various change point detection techniques, we use the "precision" and "recall" in function of a delay s as it was introduced in [1]

$$\text{Precision}(s) = \frac{1}{N_{found}} \sum_i \delta \left(\min_j \left| t_i^{found} - t_j^{known} \right| \leq s \right) \tag{7.17}$$

$$\text{Recall}(s) = \frac{1}{N_{known}} \sum_j \delta \left(\min_i \left| t_i^{found} - t_j^{known} \right| \leq s \right), \tag{7.18}$$

where N_{found} (N_{known}) is the total number of detected ("empirical") change points. The precision is the fraction of detected change points t_i^{found} that have

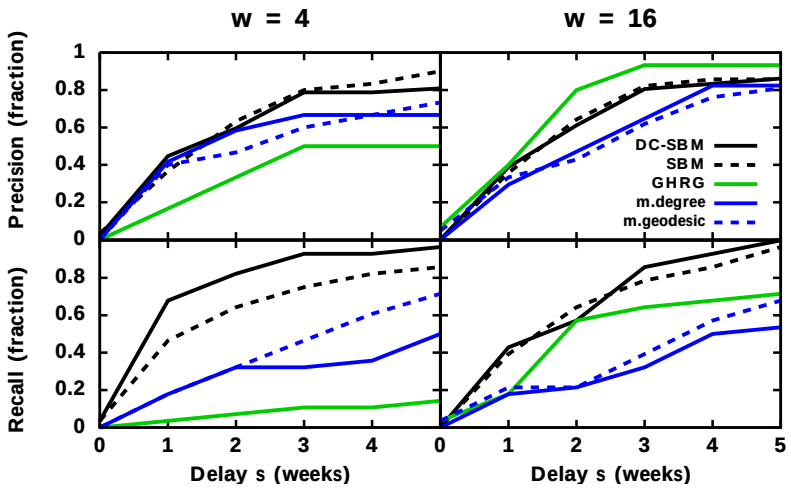


Figure 7.6: The computed precision (top) and recall (bottom) for the Enron e-mail network. Results are shown for window sizes of 4 (left) and 16 weeks (right) and for five change point detection methods.

an "empirical" event t_i^{known} within a time range of s. The recall is the fraction of "empirical" events that have a detected change point within a time range of s.

Figures 7.5, 7.7 and 7.9 show the "empirical" and the detected change points for the Enron, MIT and trade networks for two different time window widths. The corresponding results for the precision and recall are contained in Figures 7.6, 7.8 and 7.10. In order to get a better feeling of the effect of the width of the sliding time window in the change point searches, for each temporal network we have been running the algorithms for a "small" width of 4 ($w = 4$) and a "larger" width of 16 ($w = 16$).

When it comes to detecting the "empirical" change points, we find that the DC-SBM method is at least equally efficient as the SBM. Furthermore, we observe a strong sensitivity of the detected change points to the value of w. For example, whereas the SBM and DC-SBM predict more change points than the GHRG for the Enron($w = 4$), Enron($w = 16$) and MIT($w = 4$) combinations, just the opposite is observed for the other three combinations. This illustrates the sensitivity of the algorithms to the choice made with regard to the value of w.

One also faces some situations where the algorithms fail to detect the "empirical" change points. In other situations the algorithms predict a high density of change points, whereas there are no direct empirical indications that point into that direction. For example, for the international trade network, the combination DC-SBM with $w = 4$ leads to many detected change points. One could argue, however, that 4 years is too small a window for dramatic changes in the interna-

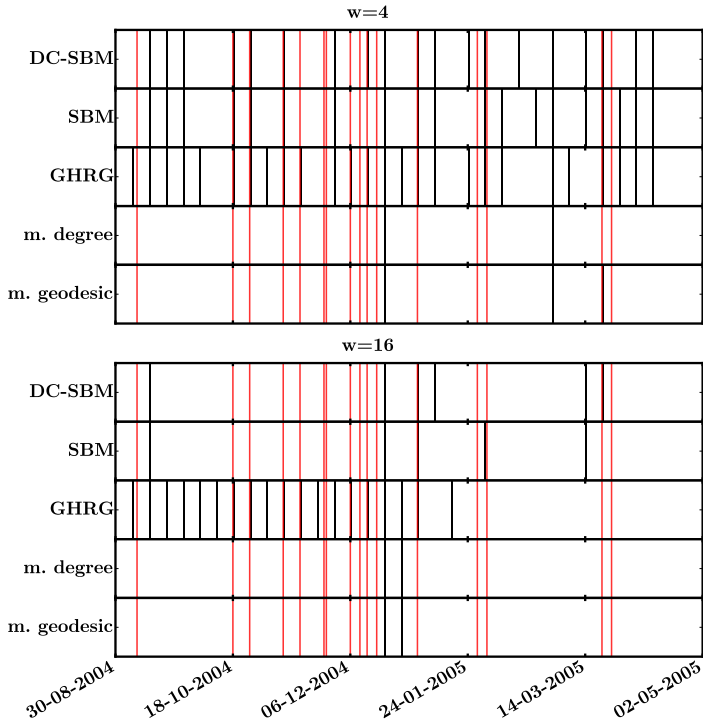


Figure 7.7: As in Figure 7.5 but for the MIT proximity network.

tional trade network to occur. For the MIT proximity network, on the other hand, all methods are performing badly for the $w = 16$ option. Here, one could argue that a time window of 4 weeks is a more natural choice to detect changes in the proximity network.

The precision of the various methods for the Enron e-mail network (Figure 7.6) is roughly the same. The GHRG method outperforms the other methods at larger window sizes. The SBM methods, in particular the DC-SBM version, perform better for the recall. The simple mean-degree and mean-geodesic methods have a decent precision but lag behind for the recall. For the precision and recall for the MIT proximity network (Figure 7.8), the GHRG method ([1]) displays a slightly better precision, but the SBM methods are slightly better at recall. Again, the simple mean-degree and mean-geodesic methods perform well for the precision but are worse for the recall. For the computed precision of the international trade network (Figure 7.10) all methods perform comparably. For the recall at $w = 4$, however, only the DC-SBM method performs better than the local methods. For a larger window size, both the SBM methods and the GHRG method perform very well.

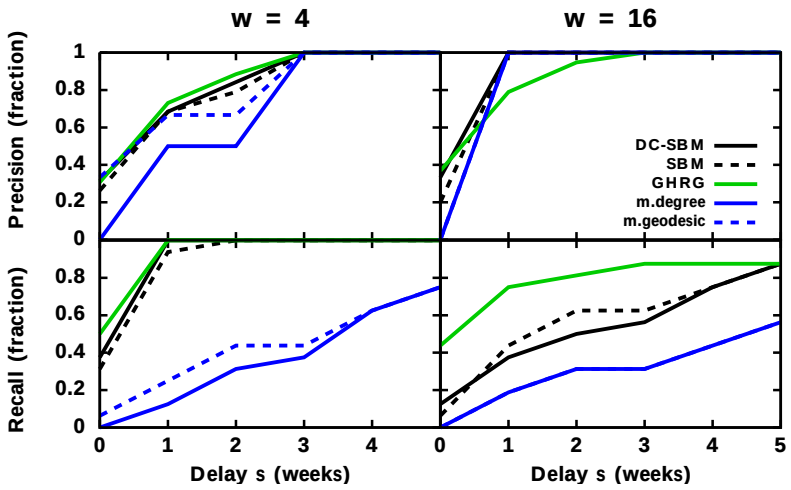


Figure 7.8: As in Figure 7.6 but for the MIT proximity network.

When comparing our results for the Enron e-mail and the MIT proximity networks with those of [1], we note some differences, especially for the Enron network. We see three possible explanations, which may together constitute a plausible explanation. Firstly, the original datasets were preprocessed in order to turn them into temporal networks. For the Enron data, a person uses several e-mail aliases, inducing uncertainties in the preprocessing of the data. Secondly, the choice of the time window width is not specified in [1] and, as shown above, the results for the change point candidates depend on that choice. Thirdly, the detected change points are sensitive to whether only active nodes or all nodes are included in the sliding time window.

7.6 Conclusion

The pioneering work of [1] developed a framework to detect change points in temporal networks based on GHRGs. In this paper we extend their methodology by adapting it to the use of SBMs as a parametric family of probability distributions for the reconstruction of empirical networks. We have made a comparative study of the detected change points on three prototypical empirical temporal networks using the GHRG and SBM based methodologies. We have done this for different sizes of the sliding time window and have also included two more simple change point detection methods in the comparison.

We find that the GHRG method and SBM methods are comparably effective in identifying the change points. In some sense, the SBM is more versatile in that it

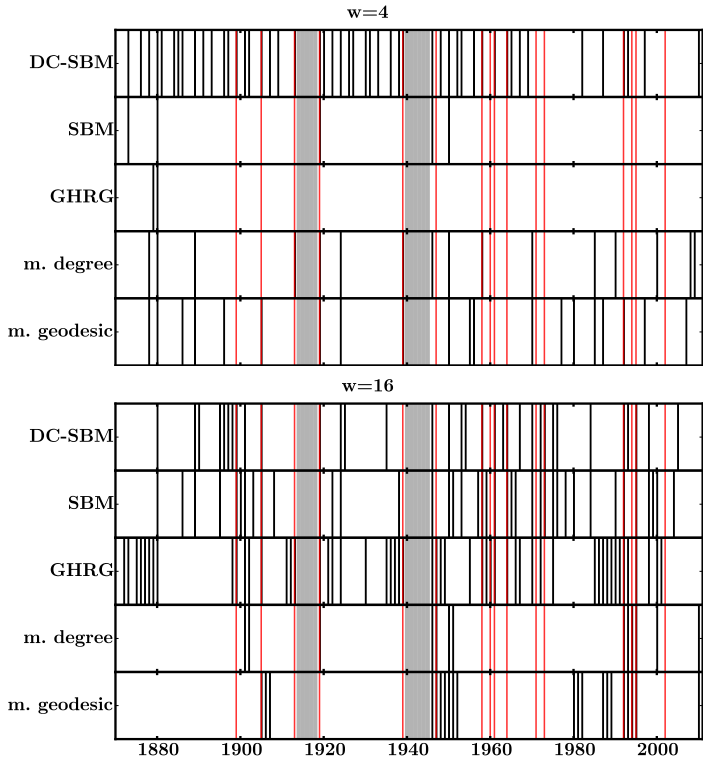


Figure 7.9: As in Figure 7.5 but for the international trade network. The widths of the sliding time windows are expressed in years.

can also deal with directed networks for example. No systematic conclusions could be drawn for the density of the detected change points. Whereas the SBM models detect more change points than the GHRG for the combinations Enron($w = 4$), Enron($w = 16$), MIT($w = 4$), just the opposite is found for the other three combinations analysed in this work. This also indicates that the choice of the size of the sliding time window affects the detected change points. When comparing the SBM and DC-SBM methodologies, the DC-SBM version has the tendency to identify a larger amount of change points. We also find some situations in which the methodologies (even dramatically) over- or under-predict the amount of "empirical" change points. Note that the SBM and GHRG are very similar models, as for an appropriate value of the number of blocks K an SBM equivalent to any GHRG can be constructed. The main difference between the two models being that the GHRG automatically determines the number of blocks at the cost of only being able to recursively partition along the block diagonal of the adjacency matrix. The SBM on the other hand can freely parametrise the full block structure

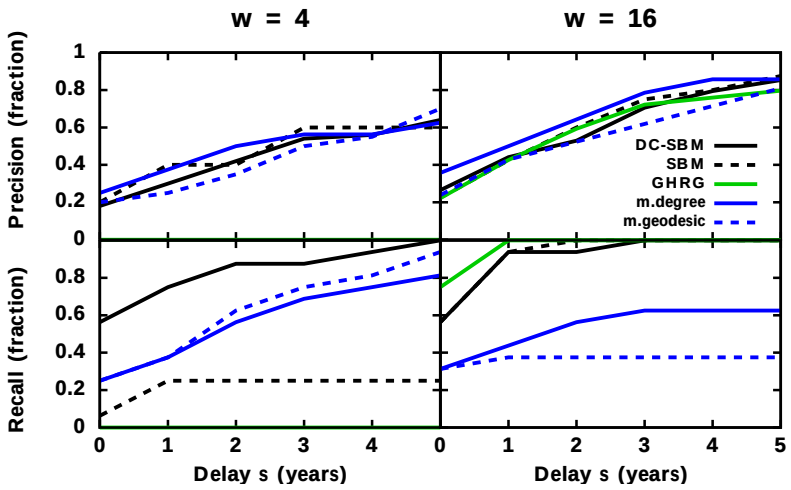


Figure 7.10: As in Figure 7.6 but for the international trade network. The widths of the sliding time windows are expressed in years.

but requires the number of blocks K to be specified. Given the similarity between SBM and GHRG it seems reasonable that they would perform similarly overall, but perform differently for different types of changes. In future work, it may be worth partitioning the problem space in more detail so that one can identify for which types of network changes the various methods perform best.

With respect to the precision and the recall, we conclude that the SBM method produces a better recall than the GHRG method, especially for sparse networks in combination with a "small" window size. The precision is only significantly outperformed by the GHRG method for one of the three studied networks. In general, the simple mean-degree and mean-geodesic methods do reasonably well for the precision but are outperformed by the sophisticated GHRG and SBM methods for the recall. This leads us to conclude that SBMs, and especially the degree-corrected SBM, are a good versatile tool for inference and analysis of complex networks. The inference of change points in temporal networks, however, is subject to some uncertainties which are connected with the adopted method and the widths of the considered sliding time windows. Methodologies based on parametric families for reconstructing the empirical networks, however, outperform the more simple methodologies.

An implementation of the proposed algorithm is available at `https://github.ugent.be/pages/sidridde/sbm_cpd`. The independence between the runs in the different time windows makes parallelisation easily attainable. In each time window, the sparse version of the belief propagation algorithm leads to a computa-

tional complexity of $\mathcal{O}((MN + N^2)K^2)$, and a memory complexity of the order $\mathcal{O}(M)$.

References

[1] Leto Peel and Aaron Clauset. *Detecting Change Points in the Large-Scale Structure of Evolving Networks*. In Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence, 2015.

[2] Mark E. J. Newman. *The structure and function of complex networks*. SIAM review, 45(2):167–256, 2003.

[3] Mark E. J. Newman, Albert-Laszlo Barabasi, and Duncan J. Watts. *The structure and dynamics of networks*. Princeton University Press, 2006.

[4] Mark E. J. Newman. *Networks: an introduction*. Oxford University Press, 2010.

[5] Phillip Bonacich. *Power and centrality: A family of measures*. AJS, 92(5):1170–1182, Mar. 1987.

[6] Stephen P. Borgatti. *Centrality and network flow*. Soc. Netw., 27(1):55–71, 2005.

[7] Stephen P. Borgatti and Martin G. Everett. *A graph-theoretic perspective on centrality*. Soc. Netw., 28(4):466–484, 2006.

[8] Pablo Moriano and Jorge Finke. *On the formation of structure in growing networks*. J. Stat. Mech., 2013(06):P06010, 2013.

[9] Santo Fortunato. *Community detection in graphs*. Phys. Rep., 486(3-5):75–174, 2010.

[10] Vincent D. Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. *Fast unfolding of communities in large networks*. J. Stat. Mech., 2008(10):P10008, 2008.

[11] Yi Chen, XL Wang, Bo Yuan, and BZ Tang. *Overlapping community detection in networks with positive and negative links*. J. Stat. Mech., 2014(3):P03021, 2014.

[12] Petter Holme. *Core-periphery organization of complex networks*. Phys. Rev. E, 72(4):046111, Oct 2005.

[13] M. Puck Rombach, Mason A. Porter, James H. Fowler, and Peter J. Mucha. *Core-periphery structure in networks*. SIAM J. Appl. Math., 74(1):167–190, 2014.

[14] Petter Holme and Jari Saramäki. *Temporal networks*. Phys. rep., 519(3):97–125, 2012.

[15] Petter Holme. *Modern temporal network theory: a colloquium*. Eur. Phys. J. B, 88(9):234, 2015.

[16] Tiago P. Peixoto. *Inferring the mesoscale structure of layered, edge-valued, and time-varying networks*. Phys. Rev. E, 92:042807, Oct 2015.

[17] Paul W. Holland, Kathryn Blackmond Laskey, and Samuel Leinhardt. *Stochastic blockmodels: First steps*. Soc. Netw., 5:109–137, June 1983.

[18] Taigo P. Peixoto. *Parsimonious Module Inference in Large Networks*. Phys. Rev. Lett., 110(14):148701, Apr 2013.

[19] Brian Karrer and Mark E. J. Newman. *Stochastic blockmodels and community structure in networks*. Phys. Rev. E, 83:016107, Jan 2011.

[20] Aurelien Decelle, Florent Krzakala, Cristopher Moore, and Lenka Zdeborová. *Asymptotic analysis of the stochastic block model for modular networks and its algorithmic applications*. Phys. Rev. E, 84:066106, Dec 2011.

[21] Xiaoran Yan, Cosma Shalizi, Jacob E Jensen, Florent Krzakala, Cristopher Moore, Lenka Zdeborová, Pan Zhang, and Yaojia Zhu. *Model selection for degree-corrected block models*. J. Stat. Mech., 2014(5):P05007, 2014.

[22] Tiago P. Peixoto. *Hierarchical Block Structures and High-Resolution Model Selection in Large Networks*. Phys. Rev. X, 4:011047, Mar 2014.

[23] Nathan Eagle, Alex (Sandy) Pentland, and David Lazer. *Inferring friendship network structure by using mobile phone data*. Proc. Natl. Acad. Sci., 106(36):15274–15278, 2009.

[24] Samuel Standaert, Stijn Ronsse, and Benjamin Vandermarliere. *Historical trade integration: globalization and the distance puzzle in the long twentieth century*. Cliometrica, 10:1–26, 2015.

8

Summary and outlook

The work presented in this dissertation can be divided into an empirical and a methodological part. Although the different chapters study a variety of economic and methodological research questions, they are bound together by a common methodological approach grounded in network theory.

In the first empirical chapter, we analyzed data on bilateral interbank exposures and bank balance sheets of the Russian interbank market from 1998 to 2004. We investigated the probability distributions of a collection of relevant interbank network characteristics. Our main observation was that all distributions are fat-tailed, and that these results were robust when considering: different time aggregates, crisis vs. non-crisis periods, and the growth vs. maturity phase of the interbank market.

The second empirical chapter studied the structure of the worldwide trade network from the 1880s to the late 1980s, looking specifically for patterns corresponding to globalization, regionalization or a core-periphery structure. Through the use of temporal stochastic block modeling, we found a strong core-periphery structure during the first wave of globalization. After WWI, however, we saw the slow dismantlement of this structure into a hub-and-spoke pattern where regional clusters are linked to a central hub, a sign of increasing regionalization.

The last empirical chapter then looked at the exchange-traded funds (ETFs) universe. Via a detailed statistical analysis of the ETF size distribution, a discrete hierarchy of sizes was discovered, which in turn led to a natural classification of the ETF universe into size layers. We found that the largest ETFs exhibit a stronger intra-layer and interlayer similarity compared with smaller ETFs. Also

the performance of the former proved significantly better than the latter.

The first two methodological chapters proposed an approach to examine the dynamical properties of transmission via novel measures on an integrated, temporally extended network representation of interaction data across time. Chapter 5 introduced the framework, provided the technical details and applied it in an exploratory way to agent-based implementations of the well-known SEIR and SEIS epidemiological models. In the follow-up Chapter 6 we further developed the previously introduced temporal knockout (TKO) score and argued that this TKO score is an effective benchmark measure for evaluating the accuracy of other, often more practical, measures of influence. Benchmarked to the traditional network measures applied to the induced flat graphs, we found that none of them are accurate predictors of network propagation influence, expressed by TKO, on the systems studied.

For the final methodological chapter we extended an existing methodology [1] for the detection of change-points in the large-scale structure of temporal networks by adding the versatile stochastic block models as generative models. This result of this extension was then compared with the original version and other, more traditional methods of change-point detection, using empirical as well as synthetic test networks. We found that none of the considered methods could consistently outperform the others when it came to detecting and locating the expected change points.

Outlook

As a final step we look ahead by discussing the ongoing and planned future research that is connected to the work presented in the preceding chapters.

In an accompanying paper to the research on ETFs we are investigating the relation between an ETF's composition compared to the market portfolio, and its performance. By comparing the weights of the stocks in a given ETF to the weights of the same stocks in the market portfolio, we arrive at a parametrization of this relation. We then check via machine learning methods whether these parameters can be used as a predictor of an ETF's performance. For now, we tentatively find that certain regions of the resulting parameter space indeed consistently over- or underperform.

In a follow-up to the work on temporal webs, the framework is now being tested on real-world interaction data. We investigate the propagation of risk factors on interbank networks, frustration on political relation networks, and emotional affect on Twitter networks. The idea is to compare the temporal web technique to the many other useful techniques out there in order to home in on the features of network propagation for which the temporal web approach provides unique insights. We are still in the exploratory phase and are momentarily focusing on the

development of computational tools.

The final ongoing project has recently been launched and will draw together all elements covered in this dissertation. The crux of the project is the exclusive access of our team to the de-identified data of five million clients of a large European bank. For a ten year time span, we have monthly records of a clients characteristics and portfolio composition. In addition we have all their financial transactions including transfers and withdrawals. This unique transaction data set constitutes a true and heretofore inaccessible treasure trove of data. The goal of the project is to unlock the full potential of this unique proprietary data set. After mapping all details of the financial transaction network and its time evolution, we will contribute to the field of "secular stagnation" and to studies of the dynamics behind income and wealth inequality. Indeed, advanced network theory applied to the data set at our disposal, provides a whole different scale, scope and time span to considerably improve on available state-of-the art studies and come up with evidence-based improved models for challenging socioeconomic issues.

References

[1] Leto Peel and Aaron Clauset. *Detecting Change Points in the Large-Scale Structure of Evolving Networks*. In Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence, 2015.

9 Samenvatting

Het werk gepresenteerd in dit proefschrift kan opgedeeld worden in een empirisch en een methodologisch stuk. Alhoewel de verschillende hoofdstukken een waaier aan economische en methodologische onderzoeksvragen aankaarten, zijn ze verbonden door een gemeenschappelijke methodologische aanpak gegrond in netwerktheorie.

In het eerste empirische hoofdstuk, analyseerden we data over bilaterale interbank leningen en balansen van banken op de Russische interbank markt in de jaren 1998 tot 2004. We onderzochten de waarschijnlijkheidsdistributies van een collectie relevante interbank netwerk karakteristieken. Onze hoofd observatie was dat al deze distributies een dikke staart hebben en dat deze resultaten robust waren bij het beschouwen van: verschillende tijds aggregaten, crisis- vs. niet-crisisperiodes, en de groei vs. de volwassen fase van de interbank markt.

Het tweede empirische hoofdstuk onderzocht de structuur van het wereldwijde handelsnetwerk van 1880 tot 1990, waarbij specifiek gezocht werd naar patronen overeenkomend met globalisatie, regionalisatie, en een kern-periferiestructuur. Door gebruik te maken van temporele *stochastic block models*, vonden we een sterke kern-periferiestructuur gedurende de eerste globaliseringsgolf. Na de Eerste Wereldoorlog, echter, zagen we een graduele overgang van deze structuur naar een *hub-and-spake* patroon waar regionale clusters gelinkt zijn met een centrale hub, een teken van toenemende regionalisatie.

Het laatste empirische hoofdstuk beschouwde het *exchange-traded funds* (ETFs) universum. Via een gedetailleerde statistische analyse van de distributie van ETF groottes, werd een discrete hiërarchie van groottes ontdekt, die op zijn beurt aan-

leiding gaf tot een natuurlijke classificatie van het ETF universum in verschillende groottelagen. We vonden dat de grootste ETFs een sterkere intra-laag en inter-laag similariteit vertoonden vergeleken met de kleinere ETFs. Daarnaast waren de prestaties van de eerstgenoemde beter dan van de laatstgenoemde.

De eerste twee methodologische hoofdstukken stelden een benadering voor om dynamische eigenschappen van propagatie te onderzoeken via een nieuwe representatie van tijdsvariërende interactie data. Hoofdstuk 5 introduceerde het raamwerk, voorzag de technische details en paste het idee toe in een verkennende manier op een agent-gebaseerde implementatie van het SEIR en SEIS epidemiologisch model. In het daarop aansluitende Hoofdstuk 6 ontwikkelden we verder de voorheen geïntroduceerde *temporal knockout* (TKO) score en argumenteerden we dat deze TKO score een effectieve maatstaf is voor het evalueren van de accuraatheid van andere, meestal meer praktische, grootheden die de invloed van een node uitdrukken. Na het vergelijken met onze maatstaf van de meer traditionele 'vlakke' netwerk grootheden, vonden we dat geen enkele van hen accuraat de TKO score voorspelde.

Voor het laatste methodologisch hoofdstuk breidden we een bestaande methodologie [1] voor het detecteren van veranderingen in de structuur van temporele netwerken uit door het toevoegen van stochastic block models als generatief model. Deze toevoeging werd dan vergeleken met de originele versie en andere, meer traditionele methodes om veranderingen te detecteren door gebruik te maken van zowel empirische als synthetische test netwerken. We vonden dat geen enkele van de beschouwde methodes consistent beter doet dan de andere wanneer het aankomt op het detecteren en lokaliseren van verwachte veranderingen.

References

[1] Leto Peel and Aaron Clauset. *Detecting Change Points in the Large-Scale Structure of Evolving Networks*. In Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence, 2015.

A

Structure and evolution of the world's historical trade patterns: Supplementary material

The definition of historical trade integration used in chapter 3 is based on that of Actual Economic Integration by [1]: *"the degree of interpenetration of economic activity among two or more countries [...] as measured at a given point in time."* The main difference is that because of data limitations, the historical trade integration index only focusses on traded goods.

Throughout this appendix, the index will be compared with other measures used in the literature. In decreasing order of availability, these are exports over total exports; exports over GDP of the sender country (e.g. [2]); the sum of exports and imports over GDP of the sender country (e.g. [3]); and the Head and Ries Index (HRI) of integration [4], which compares the bilateral trade flows with the level of internal trade of both countries.[1]

A.1 Indicators of trade integration

To measure the level of trade integration between countries we construct four measures that indicate the importance of the bilateral trade flows for the sender country.

[1] $\sqrt{X_{ij}X_{ji}/\left(X_{ii}X_{jj}\right)}$, with X_{ij} the exports from i to j and X_{ii} the internal trade in country i. Internal trade is usually approximated by subtracting exports from GDP, even though this can cause negative values for small open economies. Alternative solutions include using tariff data [5].

In order to correct for differences in scale the trade flows are normalized, since for example the importance of a million dollars worth of imports will be starkly different in the case of Latvia as opposed to the United States. Defining $X_{ij,t}$ as the total exports from the sender i to target country j in year t and $M_{ij,t}$ as the total imports from target j to sender i in year t, these measures are:

$$y_{ij,t} \equiv \{\frac{X_{ij,t}}{\sum_j X_{ij,t}},\ \frac{M_{ij,t}}{\sum_j M_{ij,t}},\ \frac{X_{ij,t}}{GDP_{i,t}},\ \frac{M_{ij,t}}{GDP_{i,t}}\}.$$

Firstly, the level of trade integration is considered high when a significant fraction of total exports go to, or imports come from, a single partner country. This normalization has the advantage that it can be computed using only trade data, but has the weakness that it does not take the overall openness to trade into account. For this reason, the last two indicators normalize import and export flows using the GDP of the sender country. However, because of the additional need for GDP data, the availability of the latter indicators is significantly lower.

To the extent that all four indicators give a similar signal the resulting index will have small confidence intervals.[2] However, when these indicators start to diverge the standard deviation will enlarge, reflecting the underlying uncertainty of the indicators. For example, in the early sixties Russia imported between one and two million dollars from Pakistan, but exported nothing. Using only exports or imports would give a very skewed view of trade relations and using the sum of both misrepresents the ambiguity of the data. Instead, they are included separately and the uncertainty of the index is used in subsequent analyses.

The historical import and export data came from three sources: the Correlates of War (COW) bilateral trade database version 3.0 [6, 7], the Research on International Commerce (RICardo) database[3] and the IMF's Direction of Trade Statistics (DoTS). Historical GDPs were provided by the Maddison project [9, 10] and supplemented with data from the Penn World Tables 8.0 [11].

In accordance with the *Real Openness* measure of [12], trade flows were measured in current, exchange rate converted, US dollars and GDP was measured in current, PPP converted, US dollars. As [12] show, using exchange rate converted GDPs (like [13]) makes the measure of openness depend on the level of the non-tradable good prices. Especially in the case of developing countries, exchange rates conversion will underestimate the GDP (the Balassa-Samuelson effect) and trade shares will be overestimated.

Following [14], the trade flows and GDPs were measured in current dollars instead of the constant 1990 US dollars (or Geary Khamis dollars), because only

[2] Since we will estimate this model using Bayesian techniques it would be more correct to use the term *highest posterior density intervals*, but for readabilitys sake, we will use *confidence interval* throughout this appendix.

[3] We are grateful to Beatrice Dedinger (`Beatrice.Dedinger@sciencespo.fr`) for providing access to the unpublised RICardo data. It was converted from pounds to US dollars using the historical exchange rate from [8].

the former can correctly compare any pair of years in the time span. In order to get GDPs in current PPP, we used the method of [13] of simply multiplying the GDPs in Geary Khamis dollars with a GDP deflator.[4] The deflator was provided by [8] and population data came from COW's National Material Capabilities database version 4.0 [15] and the Penn World Tables. More information on each of the sources and how they were converted can be found in the appendix.

This data is collected for 225 countries and territories from 1870 up to 2011, giving us a total of more than 1.8 million observations. Because trade data was significantly reduced during the World Wars, these periods were left out. It should be noted that as a lot of countries (politically speaking) did not exist at the beginning of the dataset, the total possible number of observations for this period is much lower than the more than 6.6 million suggested by the total number of countries.[5]

By including the DOTS and RICardo trade data, many colonial countries are now covered before their independence. The dataset covers colonial trade flows from as early 1880. Almost half a million of the 1.8 million observations involve a colony and three percent covers trade flows between two colonial countries. While the majority of these trade flows concern the period after World War II, a large number of colonies are covered early on.[6] Finally, it should be mentioned that the trade data only captures the official trade flows between countries. If all trade between two countries passes through a third country (re-exportation) or is smuggled, this will not be captured using this dataset.

A.2 The state-space model

Following the methodology outlined in [16], the four indicators were combined into the historical trade integration (*hti*) index using the following state-space model:

$$y_{ij,t} = C + Z * hti_{ij,t} + \epsilon_{ij,t} \quad \text{(A.1)}$$

$$hti_{ij,t} = T_t * hti_{ij,t-1} + \nu_{ij,t} \quad \text{(A.2)}$$

$$\epsilon_{ij,t} \sim N(0, H) \quad \text{(A.3)}$$

$$\nu_{ij,t} \sim N(0, Q) \quad \text{(A.4)}$$

The measurement equation (A.1) states that the four indicators $y_{ij,t}$ try to measure the level of trade integration between sender i and target country j at time t. Un-

[4] [14] starts with current, exchange rate converted, GDPs and uses the *shortcut method* to compute the current, PPP converted, GDPs. [13] on the other hand start with Maddison's GDPs in constant, PPP converted, 1990 US dollars and transforms it using a GDP deflator in current US dollars. They subsequently transform this series into current, exchange rate converted, US dollars using a similar (but inverted) *shortcut method*.

[5] n^o of countries $\times$ (n^o of countries -1) $\times$ n^o of years (excl. World Wars) = $225 \times 224 \times (2011 - 1870 + 1 - 5 - 6)$.

[6] The Overseas Countries and Territories account for the remaining colonies after the year 2000.

like for example a simple average the scaling parameters Z and C, the slope and intercept, can vary for each indicator of trade integration. Similarly, the variance of the error term ϵ can differ over all indicators, in contrast to a principle component analysis where this is kept constant. On the other hand, cross-correlation between the error terms of different indicators is ruled out: $E[\epsilon^{(k)}, \epsilon^{(m)}] = 0, \ \forall \ k \neq m$.

The state equation (A.2) allows for the trade index to depend on its previous values in the manner of an AR(1) model. This level of dependence (T_t) is assumed to be the same for all dyads. Allowing it to be different for each country couple adds more than a hundred thousand parameters to the model and slows the regression algorithm down to an infeasible degree.[7] By defining the state equation as an AR(1) process, we implicitly restrict T_t to the [-1,1] interval, including the boundary values. In other words, both stationary and non-stationary values of the *hti* index are allowed but explosive series are not.

Because the World Wars were likely to have altered trade relations significantly they were modeled as a structural break. The level of trade integration before and after each World War was assumed to be uncorrelated and the parameter of time dependence can differ over the three periods (equation A.5). In this way, the estimation of trade integration before World War I is unaffected by whatever changes happened during the interbellum or after World War II, and vice versa.

$$T_t = \begin{cases} T_1 & \text{if} \quad 1914 > t \\ T_2 & \text{if} \quad 1918 < t < 1940 \\ T_3 & \text{if} \quad 1945 < t \end{cases} \tag{A.5}$$

The issue of incomplete and missing observations is solved by replacing them with information which is entirely uncertain and does not influence the resulting index: $y = 0$, $var(\epsilon) = \infty$. This allows the model to run uninterruptedly without fundamentally changing the nature of missing data. This, in combination with the time dependency, enables us to increase the number of countries and years for which the index can be calculated without having to impute or otherwise manipulate the data [17, 18]. This matters especially for those observations where there is only partial information, for example when GDP data is missing. Without this solution for missing observations, either the index cannot be computed for those years (reducing the dataset by more than 20%), or the resulting indicator runs the risk of being distorted. The state-space model on the other hand, can still produce an estimate but will adjust the confidence intervals to reflect the lack of a complete dataset.

This model is estimated using a Bayesian Gibbs sampler algorithm, mainly because of the convenience the Gibbs sampling algorithm provides. This algorithm allows us to split up the computation of a complex (posterior) probability into

[7] Initial tests found that the time-dependency is the same for the vast majority of country couples: 94.4% of the time T_{ij} is not significantly different at the 1% level from T_{jl} with $ij \neq jl$.

several much simpler conditional probabilities. For example, if the *hti* index values were known, the state and measurement equations become very simple linear regressions models. The appendix provides more information on the estimation procedure and an excellent and detailed introduction to Bayesian Gibbs sampling and state space model can be found in chapters 7 and 8 of [17].

The Gibbs sampler ran for 6000 iterations of which the first 4000 were discarded as burn-in.[8] The remaining were used to reconstruct the posterior distribution of the level of trade integration of each dyad in each year. The resulting index is a continuous variable with values between -17 and 172. This level in itself has little meaning and can be rescaled as needed as long as relative differences over time and between countries are preserved. A higher index value corresponds to a high level of trade integration.

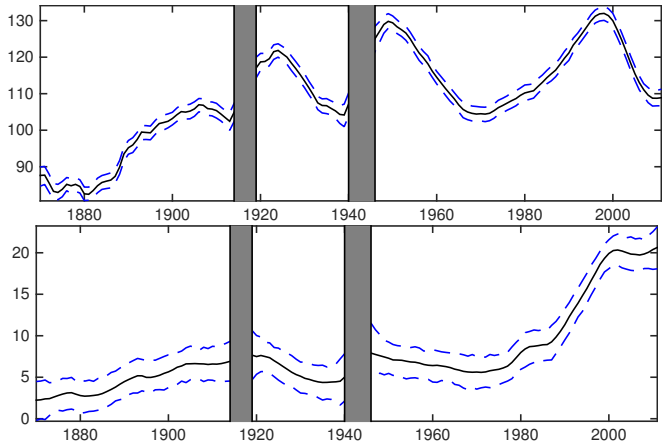
A.3 The historical trade integration index

By way of illustration, figure A.1 shows the index values for USA-Mexican bilateral trade from the perspective Mexico (panel a) and the USA (panel b). The index is standardized such that the average value and standard deviation over all countries and all years are zero and one, respectively.[9] It plots both the expected value of the index as well as its 95% confidence interval. It should be clear from this graph that the level and evolution of trade integration can differ significantly depending on the point of reference. The Mexican-US trade is highly important to the former as its index value lie entirely within the top 1 percentile. From the perspective of the US on the other hand, trade with Mexico only really becomes important from the mid-twentieth century onwards. The divergence in the evolution of the hti index values of both countries in the 21^{st} century exemplifies the fact that hti measures relative trade integration. Trade between the US and Mexico did not decrease after 2000, but trade between China (and to a lesser extent Canada) and Mexico did increase significantly. This led to a drop in the Mexico-US index, but had no effect on the US-Mexico hti index values. Furthermore, the widening of the confidence interval immediately after the World Wars illustrates the effect of a decrease in data availability in this period.

The most notable difference between the hti index and the other indicators of trade integration is the increase in data availability, especially when compared to the Head-Ries index. When using one of the alternative indicators, overall data

[8]The size of the dataset required the use of the resources of the Flemish Supercomputer Center, which was kindly provided by Ghent University, the Flemish Supercomputer Center (VSC), the Hercules Foundation and the Flemish Government − department EWI.

[9] $hti^{\star}_{ij,t} = (hti_{ij,t} - \mu)/\sigma$. With $\mu = \frac{\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n}\sum_{t=1}^{T}(hti_{ij,t})}{n(n-1)T}$, and $\sigma^2 = \frac{\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n}\sum_{t=1}^{T}(hti_{ij,t}-\mu)^2}{n(n-1)T-1}$



The index was normalised such that the mean and standard deviation for all dyads and years is respectively zero and one.

Figure A.1: The normalized historical trade index for Mexico-USA (top panel) and USA-Mexico (bottom panel).

availability decreases with 13% in the case of exports over total exports and even 38% (close to half a million dyads) when using HRI.

References

[1] Francesco Paolo Mongelli, Ettore Dorrucci, and Itai Agur. *What Does European Institutional Integration Tell Us about Trade Integration*. European Central Bank Occasional Paper Series, (40), December 2005.

[2] Giorgio Fagiolo, Javier Reyes, and Stefano Schiavo. *On the topological properties of the world trade web: A weighted network analysis*. Physica A: Statistical Mechanics and its Applications, 387(15):3868–3873, 2008.

[3] Iván Arribas, Francisco Pérez, and Emili Tortosa-Ausina. *A new interpretation of the distance puzzle based on geographic neutrality*. Economic Geography, 87(3):335–362, 2011.

[4] Keith Head and John Ries. *The erosion of colonial trade linkages after independence*. American Economic Review, 91(4):858–876, 2001.

[5] Keith Head and Thierry Mayer. *Gravity Equations: Workhorse, Toolkit, and Cookbook*. Centre for Economic Policy Research, 9322, 2013.

[6] Katherine Barbieri, Omar Keshk, and Brian Pollins. *Trading Data: Evaluating our Assumptions and Coding Rules*. Conflict Management and Peace Science, 26(4):471–491, 2009.

[7] Katherine Barbieri and Omar Keshk. *Correlates of War Project Trade Data Set Codebook, verion 3.0*, 2012.

[8] Samuel H. Williamson. *What was the U.S. GDP then?* MeasuringWorth, 2015.

[9] Madisson. *The Maddison-Project*, 2013.

[10] J. Bolt and J.L. van Zanden. *The First Update of the Maddison Project; Re-Estimating Growth Before 1820*. Maddison Project Working Paper 4, 2013.

[11] Robert C. Feenstra, Robert Inklaar, and Marcel P. Timmer. *The next Generation of the Penn World Table*, 2013.

[12] Francisco Alcalá and Antonio Ciccone. *Trade and Productivity*. The Quarterly Journal of Economics, 119(2):613–646, 2004.

[13] Mariko J. Klasing and Petros Milionis. *Quantifying the Evolution of World Trade, 1870-1949*. Journal of International Economics, 92(1):185–197, 2014.

[14] Leandro Prados de la Escosura. *International Comparisons of Real Product, 1820-1990: And alternative Data Set*. Explorations in Economic History, 37:1–41, 2000.

[15] J. David Singer, Stuart Bremer, and John Stuckey. *Capability Distribution, Uncertainty, and Major Power War, 1820-1965*. In Bruce Russet, editor, Peace, War, and Numbers, pages 19–48. Sage, Beverly Hills, 1972.

[16] Glenn Rayp and Samuel Standaert. *Measuring Actual Integration: An Outline of a Bayesian State-Space Approach*. In Philippe De Lombaerde and E. Saucedo, editors, Indicator-Based Monitoring of Regional Economic Integration, UNU series on regionalism. Springer, Dordrecht-New York, forthcoming.

[17] Chang-Jin Kim and Charles R. Nelson. *State-Space Models with Regime Switching: Classical and Gibbs-Sampling Approaches with Applications*. MIT Press, 1999.

[18] J. Durbin and S.J. Koopman. *Time Series Analysis by State Space Methods*. Oxford University Press, 2 edition, 2012.

B

Benchmarking measures of network influence: Supplementary material

B.1 Outline

Model Scenarios and Infection Sizes: Includes a 3D histogram with a row for each scenario showing the frequency of infections of each size. We also have a set of twelve 3D histograms (one for each scenario) with a row for each of the 25 skeletons showing the disease variation resulting from network structure; however, in consideration of space and the real focus of this paper they are excluded (available upon request). We also provide a table of the mean and standard deviations of the raw magnitudes for each scenario. When excluding the duds the distributions approximate normal distributions, but the large numbers of duds make the normal approximation inappropriate and it is *not* the case that the disease results follow any single distribution with the mean and standard deviations in the table. The mean and standard deviation do, however, capture the relative all-things-considered expected infection sizes for each scenario.

Correlations between Agent-Initialized Magnitude and TKO Measures: The Pearson and Spearman correlation coefficients between (a) the disease magnitude reached when agent i is the initial agent and (b) four different versions of the TKO aggregated across time for agent i. The correlations are performed separately for each scenario between the lists of values for all 200 agents combined across all 25

network skeletons. With the highest scores near 0.50 and most much lower, the result is that measuring an agent's impact using the super-spreader approach alone is not accurate in capturing an agent's actual influence compared to TKO.

Comparisons of Network Measures to TKO scores: This section starts with one page of further methodological description, especially about the flattened observed interaction dynamics networks. Following that are eight pages of table triplets each showing the Pearson, Spearman, and Top Ten comparisons between each of five common network centrality measures. Each of the four TKO variation has it's one page of tables for both the base and the flattened networks. Because the base and unweighted flattened networks are nearly identical, so are the correlations. The weighted versions of the flattened measure are excluded in consideration of the space to describe them in light of the result that they also do not significantly covary with any TKO measures.

B.2 Model Scenarios and Infection Sizes

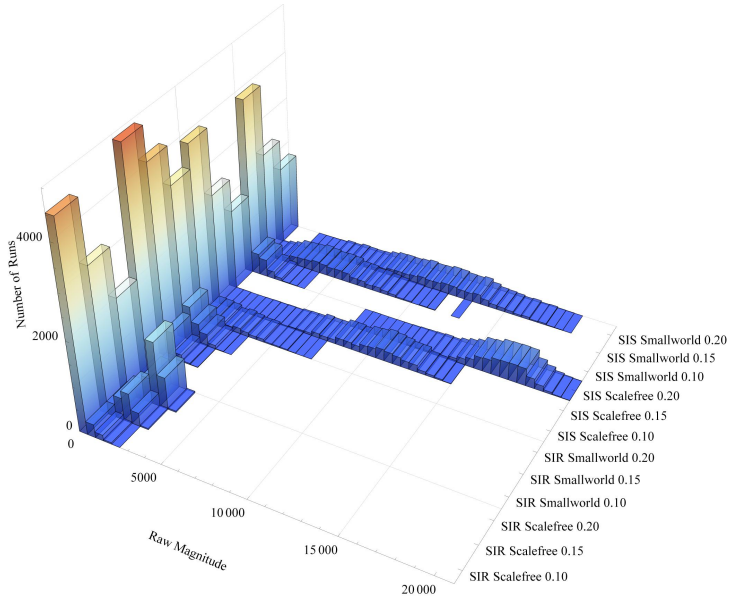


Figure B.1: Results histogram of infection spread in terms of the number of temporal nodes infected (raw magnitude) across 5000 runs for each scenario (one run initialized at each of 200 agents for each of the 25 base network implementations). Notice that a very large proportion of runs are "duds" in which the infection fails to spread beyond 50 temporal nodes. The SIS models naturally have greater magnitude values due to reinfection. These dynamics are typical of SIR and SIS models with similar parameters.

Infection Type	Network Type	Infection Probability	Mean Magnitude	Magnitude StDev	Percent Duds
SIR	Scale Free	0.10	143.352	288.549	0.625
SIR	Scale Free	0.15	584.744	774.628	0.482
SIR	Scale Free	0.20	1296.44	1142.24	0.380
SIR	Small World	0.10	88.9266	131.743	0.584
SIR	Small World	0.15	227.321	324.207	0.457
SIR	Small World	0.20	445.033	559.017	0.352
SIS	Scale Free	0.10	548.746	1155.19	0.593
SIS	Scale Free	0.15	5003.03	5237.44	0.445
SIS	Scale Free	0.20	10800.6	8150.76	0.344
SIS	Small World	0.10	308.734	536.106	0.557
SIS	Small World	0.15	2526.97	2839.26	0.433
SIS	Small World	0.20	7036.79	5623.18	0.333

Table B.1: Results summary of infection spread for each model variation. Each row aggregates 5000 runs (one run initialized at each of 200 agents for each of the 25 base network implementations). Duds are defined as runs in which the raw magnitude is fewer than 50 agent-times.

B.3 Correlations between Agent-Initialized Magnitude and TKO Measures

Pearson Correlations of Agent-Initialized Magnitude and TKO Measures.

Disease Type	Network Type	InfectionRate	MaxProportion	MaxDeltaFraction	AveProportion	AveDeltaFraction
SIR	scalefree	0.10	0.402831	0.404905	0.288126	0.292081
SIR	scalefree	0.15	0.0674457	0.246309	0.0644679	0.157114
SIR	scalefree	0.20	0.0457449	0.219299	0.0781154	0.158489
SIR	smallworld	0.10	0.494424*	0.471569	0.366466	0.363984
SIR	smallworld	0.15	0.043098	0.264589	0.0770296	0.188738
SIR	smallworld	0.20	0.0297556	0.19209	0.0154106	0.118696
SIS	scalefree	0.10	0.346553	0.375933	0.268197	0.282026
SIS	scalefree	0.15	0.0566606	0.247974	0.0838702	0.153358
SIS	scalefree	0.20	0.0446214	0.233998	0.0585834	0.107849
SIS	smallworld	0.10	0.404204	0.417726	0.353329	0.370916
SIS	smallworld	0.15	0.0240389	0.201376	0.0452133	0.149954
SIS	smallworld	0.20	0.0502245	0.190489	0.0420603	0.10682

Table B.2: The mean Pearson correlations coefficients of (a) the disease magnitude given an agent is the initially infected agent and (b) the TKO score for that agent. The on-average low correlations imply that using the disease spread based on initial infection is a poor measure of influence. Furthermore, the correlations are nearly always worse with increasing infection rates (and hence increasing magnitudes and fewer dud runs) implying that much of the ability to match TKO relies on the cases in which both scores are near zero.

Spearman Correlations of Agent-Initialized Magnitude and TKO Measures.

Disease Type	Network Type	InfectionRate	MaxProportion	MaxDeltaFraction	AveProportion	AveDeltaFraction
SIR	scalefree	0.10	0.40153	0.305474	0.388263	0.308888
SIR	scalefree	0.15	0.0630928	0.221007	0.105923	0.193733
SIR	scalefree	0.20	0.0277719	0.203116	0.0667476	0.168757
SIR	smallworld	0.10	0.516584*	0.420064	0.461123	0.404163
SIR	smallworld	0.15	0.0671833	0.292652	0.149355	0.270376
SIR	smallworld	0.20	0.0375412	0.222854	0.0577158	0.171541
SIS	scalefree	0.10	0.296555	0.252962	0.277245	0.235637
SIS	scalefree	0.15	0.0487628	0.198384	0.0806729	0.13593
SIS	scalefree	0.20	0.0370982	0.184936	0.027529	0.069931
SIS	smallworld	0.10	0.335403	0.280811	0.29627	0.255873
SIS	smallworld	0.15	0.0256655	0.190925	0.0797812	0.160578
SIS	smallworld	0.20	0.0449195	0.187943	0.0447358	0.11443

Table B.3: The mean Spearman Rank correlation coefficients (rho) of (a) the disease magnitude given an agent is the initially infected agent and (b) the TKO score for that agent. The correlations reveal similar values and a similar pattern to the Pearson correlations, reinforcing that using the disease spread based on initial infection is a poor measure of influence.

B.4 Comparisons of Network Measures to TKO scores

The following twelve sets of three data tables present the results of determining how well common network centrality measures capture agent influence. Although the main result is that none of the network measures successfully capture/predict agent influence as measured by four versions of TKO in any scenario, the specific changes in the data reveal patterns – and those patterns may point to improved measures.

Although the paper focuses on the base network analysis, we also analyzed the network generated by flattening the observed interactions. We record who interacts with whom over time in the temporal network skeleton, then we flatten this skeleton to achieve both a weighted by interaction frequency and an unweighted flat network representation. If the model runs long enough the observed interactions converge to the base network of potential interactions, but in many applications the flattened network is observable/derivable from data while the base network is unknown and/or theoretical. In our simulations, because the probability that a given link is active in a time step is $\propto 1/k$, k is low (typically single digit except a few agents in the scale free networks), and there are 200 time steps, the base and unweighted flattened graphs are nearly identical.

Because for each base network we generate the skeleton including all transition and interaction probabilities, the empirically derived flattened network connections are always the same for each run of the same skeleton (i.e., starting from each agent). In the current model the infection state does not alter the interaction probability. If it did, then the observed transitions would vary from run to run even using the same network skeleton because what is stored in the skeleton is a set of draws from probability distributions rather than a fixed interaction structure. If, for example, being infectious reduced the probability of interaction, then the probability stored in the skeleton would be compared to a different interaction threshold and thus could alter which interactions occur. However, using the same skeletons for multiple runs of different dynamics on the same structure at least satisfies the Markov condition for these simulations, which is not maintained when running the dynamics independently for each initial agent run.

Flattened graphs are potentially better at tracking influence because they allow one to create weighted networks from the observed interaction frequencies. However, in our experiments the correlation valued between TKO and the weighted network centrality measures were no better, although they were slightly different. For this reason and considerations of space we have excluded them from this paper.

Maximum TKO and Base Interaction Network

Pearson Correlations of Centrality Measures and Maximum TKO on Base Network

Disease Type	Network Type	InfectionRate	Degree	Closeness	Betweenness	Eigenvector	Katz
SIR	scalefree	0.10	0.12254	0.09622	0.09557	0.08855	0.06909
SIR	scalefree	0.15	0.1475	0.11315	0.127	0.11531	0.09876
SIR	scalefree	0.20	0.14354	0.11432	0.11925	0.11355	0.09922
SIR	smallworld	0.10	0.00727	0.0654	0.03754	0.00047	-0.0051
SIR	smallworld	0.15	0.05936	0.0854	0.11007	0.08465	0.0749
SIR	smallworld	0.20	0.08282	0.17901	0.1399	0.03241	0.07405
SIS	scalefree	0.10	0.13835	0.09571	0.11868	0.09935	0.07816
SIS	scalefree	0.15	0.11807	0.08243	0.09698	0.0866	0.07341
SIS	scalefree	0.20	0.09523	0.07743	0.07644	0.07262	0.06029
SIS	smallworld	0.10	0.02154	0.07993	0.03546	-0.01334	0.00489
SIS	smallworld	0.15	0.0637	0.11582	0.11384	0.02679	0.05537
SIS	smallworld	0.20	0.11755	0.13727	0.22958	0.06551	0.13365

Spearman Correlations of Centrality Measures and Maximum TKO on Base Network

Disease Type	Network Type	InfectionRate	Degree	Closeness	Betweenness	Eigenvector	Katz
SIR	scalefree	0.10	0.19228	0.10423	0.18884	0.09346	0.06062
SIR	scalefree	0.15	0.19853	0.10738	0.17779	0.10471	0.07401
SIR	scalefree	0.20	0.21217	0.12175	0.20802	0.11573	0.08827
SIR	smallworld	0.10	0.01237	0.02756	0.03768	0.03103	0.01127
SIR	smallworld	0.15	0.05613	0.09342	0.09055	0.06225	0.04637
SIR	smallworld	0.20	0.08772	0.17172	0.10234	0.07409	0.08572
SIS	scalefree	0.10	0.18963	0.0922	0.18953	0.08388	0.0518
SIS	scalefree	0.15	0.18795	0.08439	0.17362	0.08274	0.05498
SIS	scalefree	0.20	0.19385	0.08551	0.19045	0.08034	0.04848
SIS	smallworld	0.10	0.01626	0.04382	0.02805	-0.00418	-0.00617
SIS	smallworld	0.15	0.08308	0.15921	0.14567	0.03365	0.05083
SIS	smallworld	0.20	0.11411	0.16911	0.20342	0.09623	0.13965

Top Ten Overlap of Centrality Measures and Maximum TKO on Base Network

Disease Type	Network Type	InfectionRate	Degree	Closeness	Betweenness	Eigenvector	Katz
SIR	scalefree	0.10	0.004	0.012	0.008	0.	0.
SIR	scalefree	0.15	0.004	0.008	0.004	0.	0.
SIR	scalefree	0.20	0.008	0.004	0.016	0.	0.
SIR	smallworld	0.10	0.	0.008	0.004	0.	0.
SIR	smallworld	0.15	0.004	0.004	0.02	0.	0.
SIR	smallworld	0.20	0.004	0.004	0.008	0.	0.
SIS	scalefree	0.10	0.008	0.012	0.012	0.	0.
SIS	scalefree	0.15	0.	0.004	0.	0.	0.
SIS	scalefree	0.20	0.008	0.012	0.004	0.	0.
SIS	smallworld	0.10	0.	0.008	0.012	0.	0.
SIS	smallworld	0.15	0.016	0.004	0.012	0.	0.
SIS	smallworld	0.20	0.02	0.008	0.028	0.	0.

Table B.4: The Pearson and Spearman correlations as well as the average percent of matching Top Ten agents between the maximum proportional TKO score with each of five base network agent centrality scores.

Maximum TKO and Flattened Interaction Network - Unweighted

Pearson Correlations of Unweighted Centrality Measures and Maximum TKO on Flattened Network

Disease Type	Network Type	InfectionRate	Degree	Closeness	Betweenness	Eigenvector	Katz
SIR	scalefree	0.10	0.12308	0.09402	0.09698	0.071	-0.01539
SIR	scalefree	0.15	0.14731	0.11036	0.1275	0.05439	0.01392
SIR	scalefree	0.20	0.14399	0.11477	0.12008	0.03457	0.00323
SIR	smallworld	0.10	0.00727	0.0654	0.03754	-0.03346	-0.0077
SIR	smallworld	0.15	0.05936	0.0854	0.11007	0.04161	-0.00132
SIR	smallworld	0.20	0.08282	0.17901	0.1399	-0.02131	0.01624
SIS	scalefree	0.10	0.13867	0.09368	0.11945	0.07499	0.00385
SIS	scalefree	0.15	0.11804	0.07891	0.09743	0.06998	0.02266
SIS	scalefree	0.20	0.09568	0.07786	0.07724	0.02772	0.01391
SIS	smallworld	0.10	0.02154	0.07993	0.03546	-0.00236	-0.00627
SIS	smallworld	0.15	0.0637	0.11582	0.11384	0.00063	0.01076
SIS	smallworld	0.20	0.11755	0.13727	0.22958	-0.03648	0.02426

Spearman Correlations of Unweighted Centrality Measures and Maximum TKO on Flattened Network

Disease Type	Network Type	InfectionRate	Degree	Closeness	Betweenness	Eigenvector	Katz
SIR	scalefree	0.10	0.19227	0.10395	0.18916	0.06804	-0.01847
SIR	scalefree	0.15	0.19851	0.10683	0.18072	0.06697	0.02382
SIR	scalefree	0.20	0.2122	0.12236	0.20727	0.03882	0.00276
SIR	smallworld	0.10	0.01237	0.02756	0.03768	-0.05264	-0.00279
SIR	smallworld	0.15	0.05613	0.09342	0.09055	0.05751	-0.00228
SIR	smallworld	0.20	0.08772	0.17172	0.10234	0.02494	0.01547
SIS	scalefree	0.10	0.18963	0.09154	0.19067	0.05854	-0.00145
SIS	scalefree	0.15	0.1879	0.08214	0.17613	0.06379	0.03914
SIS	scalefree	0.20	0.19386	0.08589	0.18989	0.02634	0.01455
SIS	smallworld	0.10	0.01626	0.04382	0.02805	-0.01907	-0.01421
SIS	smallworld	0.15	0.08308	0.15921	0.14567	0.04287	0.00706
SIS	smallworld	0.20	0.11411	0.16911	0.20342	0.0116	-0.0002

Top Ten Overlap of Unweighted Centrality Measures and Maximum TKO on Flattened Network

Disease Type	Network Type	InfectionRate	Degree	Closeness	Betweenness	Eigenvector	Katz
SIR	scalefree	0.10	0.004	0.004	0.004	0.	0.
SIR	scalefree	0.15	0.	0.008	0.012	0.	0.
SIR	scalefree	0.20	0.008	0.	0.016	0.	0.
SIR	smallworld	0.10	0.	0.008	0.004	0.	0.
SIR	smallworld	0.15	0.004	0.004	0.02	0.	0.
SIR	smallworld	0.20	0.004	0.004	0.008	0.	0.
SIS	scalefree	0.10	0.008	0.016	0.012	0.	0.
SIS	scalefree	0.15	0.004	0.004	0.	0.	0.
SIS	scalefree	0.20	0.008	0.012	0.004	0.	0.
SIS	smallworld	0.10	0.	0.008	0.012	0.	0.
SIS	smallworld	0.15	0.016	0.004	0.012	0.	0.
SIS	smallworld	0.20	0.02	0.008	0.028	0.	0.

Table B.5: The Pearson and Spearman correlations as well as the average percent of matching Top Ten agents between the maximum proportional TKO score with each of five flattened observed interaction network agent centrality scores.

Maximum Delta Fraction TKO and Base Interaction Network

Pearson Correlations of Centrality Measures and Maximum Delta TKO on Base Network

Disease Type	Network Type	InfectionRate	Degree	Closeness	Betweenness	Eigenvector	Katz
SIR	scalefree	0.10	0.10286	0.08344	0.08	0.07491	0.05781
SIR	scalefree	0.15	0.16692	0.13195	0.15128	0.13926	0.12469
SIR	scalefree	0.20	0.15331	0.12103	0.12757	0.1196	0.10376
SIR	smallworld	0.10	0.00478	0.08652	0.04694	-0.00663	-0.00979
SIR	smallworld	0.15	0.08566	0.1837	0.14683	0.07832	0.09028
SIR	smallworld	0.20	0.11087	0.20727	0.16839	0.04081	0.10154
SIS	scalefree	0.10	0.11816	0.08506	0.10179	0.08919	0.0728
SIS	scalefree	0.15	0.13843	0.10067	0.11605	0.1063	0.09209
SIS	scalefree	0.20	0.11433	0.08954	0.09371	0.08694	0.07255
SIS	smallworld	0.10	0.01778	0.11559	0.04528	-0.00982	0.00079
SIS	smallworld	0.15	0.08102	0.23611	0.15822	0.04475	0.07326
SIS	smallworld	0.20	0.14115	0.20109	0.26869	0.07395	0.15273

Spearman Correlations of Centrality Measures and Maximum Delta TKO on Base Network

Disease Type	Network Type	InfectionRate	Degree	Closeness	Betweenness	Eigenvector	Katz
SIR	scalefree	0.10	0.19207	0.10484	0.19034	0.09269	0.05831
SIR	scalefree	0.15	0.22213	0.12246	0.21438	0.11981	0.08953
SIR	scalefree	0.20	0.23221	0.13404	0.2258	0.1246	0.09275
SIR	smallworld	0.10	0.01617	0.05484	0.0493	0.0346	0.00932
SIR	smallworld	0.15	0.06147	0.17675	0.09661	0.07697	0.06284
SIR	smallworld	0.20	0.09899	0.24107	0.14145	0.07399	0.08672
SIS	scalefree	0.10	0.18397	0.08878	0.18672	0.07923	0.04705
SIS	scalefree	0.15	0.22487	0.09828	0.21779	0.0964	0.06375
SIS	scalefree	0.20	0.22593	0.09769	0.22199	0.08987	0.05166
SIS	smallworld	0.10	0.01798	0.07497	0.04538	-0.01973	-0.0145
SIS	smallworld	0.15	0.08985	0.26058	0.16032	0.05387	0.06764
SIS	smallworld	0.20	0.13121	0.23591	0.23906	0.09482	0.14568

Top Ten Overlap of Centrality Measures and Maximum Delta TKO on Base Network

Disease Type	Network Type	InfectionRate	Degree	Closeness	Betweenness	Eigenvector	Katz
SIR	scalefree	0.10	0.008	0.012	0.008	0.	0.
SIR	scalefree	0.15	0.016	0.004	0.016	0.	0.
SIR	scalefree	0.20	0.004	0.004	0.012	0.	0.
SIR	smallworld	0.10	0.004	0.008	0.004	0.	0.
SIR	smallworld	0.15	0.02	0.008	0.024	0.	0.
SIR	smallworld	0.20	0.004	0.008	0.02	0.	0.
SIS	scalefree	0.10	0.016	0.004	0.016	0.	0.
SIS	scalefree	0.15	0.008	0.004	0.004	0.	0.
SIS	scalefree	0.20	0.008	0.016	0.008	0.	0.
SIS	smallworld	0.10	0.008	0.016	0.016	0.	0.
SIS	smallworld	0.15	0.008	0.004	0.024	0.	0.
SIS	smallworld	0.20	0.028	0.012	0.024	0.	0.

Table B.6: The Pearson and Spearman correlations as well as the average percent of matching Top Ten agents between the maximum change in fractional TKO score with each of five base network agent centrality scores.

Maximum Delta Fraction TKO and Flattened Interaction Network - Unweighted

Pearson Correlations of Unweighted Centrality Measures and Maximum Delta TKO on Flattened Network

Disease Type	Network Type	InfectionRate	Degree	Closeness	Betweenness	Eigenvector	Katz
SIR	scalefree	0.10	0.10322	0.08126	0.08119	0.0504	-0.00984
SIR	scalefree	0.15	0.16637	0.12859	0.15112	0.06161	0.02626
SIR	scalefree	0.20	0.15391	0.12173	0.12867	0.04181	0.00983
SIR	smallworld	0.10	0.00478	0.08652	0.04694	-0.03715	-0.00275
SIR	smallworld	0.15	0.08566	0.1837	0.14683	0.07864	0.00211
SIR	smallworld	0.20	0.11087	0.20727	0.16839	-0.0042	-0.00451
SIS	scalefree	0.10	0.1181	0.08238	0.102	0.06581	0.00915
SIS	scalefree	0.15	0.13813	0.09611	0.11633	0.08351	0.02489
SIS	scalefree	0.20	0.11505	0.09018	0.09494	0.03574	0.01906
SIS	smallworld	0.10	0.01778	0.11559	0.04528	-0.01955	0.00412
SIS	smallworld	0.15	0.08102	0.23611	0.15822	0.03885	0.01473
SIS	smallworld	0.20	0.14115	0.20109	0.26869	-0.04363	0.01438

Spearman Correlations of Unweighted Centrality Measures and Maximum Delta TKO on Flattened Network

Disease Type	Network Type	InfectionRate	Degree	Closeness	Betweenness	Eigenvector	Katz
SIR	scalefree	0.10	0.19206	0.10443	0.19097	0.06511	-0.02302
SIR	scalefree	0.15	0.22211	0.12175	0.21708	0.07393	0.01792
SIR	scalefree	0.20	0.23223	0.13514	0.22504	0.05252	0.00736
SIR	smallworld	0.10	0.01617	0.05484	0.0493	-0.05967	0.00591
SIR	smallworld	0.15	0.06147	0.17675	0.09661	0.11312	0.00537
SIR	smallworld	0.20	0.09899	0.24107	0.14145	0.02893	0.0099
SIS	scalefree	0.10	0.18396	0.08785	0.18809	0.05192	0.00408
SIS	scalefree	0.15	0.22485	0.09573	0.22064	0.0759	0.03789
SIS	scalefree	0.20	0.22593	0.09808	0.22135	0.04268	0.02456
SIS	smallworld	0.10	0.01798	0.07497	0.04538	-0.01514	-0.01234
SIS	smallworld	0.15	0.08985	0.26058	0.16032	0.10784	0.0099
SIS	smallworld	0.20	0.13121	0.23591	0.23906	0.00745	-0.01646

Top Ten Overlap of Unweighted Centrality Measures and Maximum Delta TKO on Flattened Network

Disease Type	Network Type	InfectionRate	Degree	Closeness	Betweenness	Eigenvector	Katz
SIR	scalefree	0.10	0.008	0.012	0.012	0.	0.
SIR	scalefree	0.15	0.012	0.004	0.016	0.	0.
SIR	scalefree	0.20	0.004	0.	0.012	0.	0.
SIR	smallworld	0.10	0.004	0.008	0.004	0.	0.
SIR	smallworld	0.15	0.02	0.008	0.024	0.	0.
SIR	smallworld	0.20	0.004	0.008	0.02	0.	0.
SIS	scalefree	0.10	0.016	0.008	0.016	0.	0.
SIS	scalefree	0.15	0.008	0.004	0.004	0.	0.
SIS	scalefree	0.20	0.012	0.008	0.012	0.	0.
SIS	smallworld	0.10	0.008	0.016	0.016	0.	0.
SIS	smallworld	0.15	0.008	0.004	0.024	0.	0.
SIS	smallworld	0.20	0.028	0.012	0.024	0.	0.

Table B.7: The Pearson and Spearman correlations as well as the average percent of matching Top Ten agents between the maximum change in fractional TKO score with each of five flattened observed interaction network agent centrality scores.

Mean TKO and Base Interaction Network

Pearson Correlations of Centrality Measures and Mean TKO on Base Network

Disease Type	Network Type	InfectionRate	Degree	Closeness	Betweenness	Eigenvector	Katz
SIR	scalefree	0.10	0.12721	0.10583	0.09908	0.09757	0.07997
SIR	scalefree	0.15	0.12514	0.09557	0.10187	0.0977	0.08398
SIR	scalefree	0.20	0.13992	0.11872	0.11503	0.11485	0.10137
SIR	smallworld	0.10	-0.00492	0.03076	0.01898	-0.00766	-0.01483
SIR	smallworld	0.15	0.04866	0.07044	0.08208	0.08047	0.06403
SIR	smallworld	0.20	0.06674	0.12695	0.10153	0.02619	0.05665
SIS	scalefree	0.10	0.13528	0.08851	0.11109	0.09088	0.06804
SIS	scalefree	0.15	0.16164	0.09822	0.127	0.10646	0.08486
SIS	scalefree	0.20	0.23309	0.16747	0.19103	0.16756	0.13834
SIS	smallworld	0.10	0.02536	0.0545	0.01982	-0.0126	0.0042
SIS	smallworld	0.15	0.06834	0.10208	0.11029	0.02545	0.05613
SIS	smallworld	0.20	0.12315	0.17406	0.24128	0.0522	0.12552

Spearman Correlations of Centrality Measures and Mean TKO on Base Network

Disease Type	Network Type	InfectionRate	Degree	Closeness	Betweenness	Eigenvector	Katz
SIR	scalefree	0.10	0.1905	0.09194	0.19137	0.08058	0.04674
SIR	scalefree	0.15	0.20983	0.11228	0.19027	0.1079	0.07657
SIR	scalefree	0.20	0.22161	0.13071	0.21508	0.12272	0.09371
SIR	smallworld	0.10	0.00607	0.01531	0.0249	0.02495	0.00735
SIR	smallworld	0.15	0.05186	0.08794	0.08288	0.07049	0.05397
SIR	smallworld	0.20	0.05635	0.15674	0.0957	0.05611	0.05825
SIS	scalefree	0.10	0.20006	0.08366	0.2032	0.07421	0.03927
SIS	scalefree	0.15	0.27654	0.10011	0.26282	0.09862	0.05359
SIS	scalefree	0.20	0.32109	0.13576	0.31428	0.12637	0.07313
SIS	smallworld	0.10	0.00761	0.02375	0.01917	-0.01435	-0.01696
SIS	smallworld	0.15	0.06719	0.15883	0.13379	0.0364	0.04759
SIS	smallworld	0.20	0.11495	0.20274	0.21729	0.0699	0.11563

Top Ten Overlap of Centrality Measures and Mean TKO on Base Network

Disease Type	Network Type	InfectionRate	Degree	Closeness	Betweenness	Eigenvector	Katz
SIR	scalefree	0.10	0.012	0.008	0.012	0.	0.
SIR	scalefree	0.15	0.012	0.02	0.	0.	0.
SIR	scalefree	0.20	0.02	0.004	0.016	0.	0.
SIR	smallworld	0.10	0.004	0.008	0.004	0.	0.
SIR	smallworld	0.15	0.008	0.024	0.008	0.	0.
SIR	smallworld	0.20	0.016	0.008	0.016	0.	0.
SIS	scalefree	0.10	0.012	0.008	0.016	0.	0.
SIS	scalefree	0.15	0.012	0.008	0.008	0.	0.
SIS	scalefree	0.20	0.016	0.028	0.02	0.	0.
SIS	smallworld	0.10	0.004	0.004	0.016	0.	0.
SIS	smallworld	0.15	0.02	0.012	0.016	0.	0.
SIS	smallworld	0.20	0.012	0.016	0.032	0.	0.

Table B.8: The Pearson and Spearman correlations as well as the average percent of matching Top Ten agents between the mean proportional TKO score with each of five base network agent centrality scores.

Mean TKO and Flattened Interaction Network - Unweighted

Pearson Correlations of Unweighted Centrality Measures and Mean TKO on Flattened Network

Disease Type	Network Type	InfectionRate	Degree	Closeness	Betweenness	Eigenvector	Katz
SIR	scalefree	0.10	0.12815	0.10446	0.10138	0.06	-0.0147
SIR	scalefree	0.15	0.12567	0.09498	0.10378	0.06054	-0.00984
SIR	scalefree	0.20	0.14023	0.11871	0.1158	0.04537	0.01406
SIR	smallworld	0.10	-0.00492	0.03076	0.01898	-0.0322	-0.01334
SIR	smallworld	0.15	0.04866	0.07044	0.08208	0.04724	-0.00565
SIR	smallworld	0.20	0.06674	0.12695	0.10153	-0.00722	0.01564
SIS	scalefree	0.10	0.13606	0.08709	0.11292	0.08012	0.01291
SIS	scalefree	0.15	0.16208	0.09486	0.12839	0.09913	0.01734
SIS	scalefree	0.20	0.23404	0.16759	0.19278	0.10685	0.01614
SIS	smallworld	0.10	0.02536	0.0545	0.01982	0.00807	0.00569
SIS	smallworld	0.15	0.06834	0.10208	0.11029	0.0242	0.01523
SIS	smallworld	0.20	0.12315	0.17406	0.24128	0.00142	-0.004

Spearman Correlations of Unweighted Centrality Measures and Mean TKO on Flattened Network

Disease Type	Network Type	InfectionRate	Degree	Closeness	Betweenness	Eigenvector	Katz
SIR	scalefree	0.10	0.19049	0.09173	0.19154	0.05292	-0.0181
SIR	scalefree	0.15	0.20984	0.11246	0.19295	0.07219	0.02852
SIR	scalefree	0.20	0.22161	0.13124	0.2144	0.04265	-0.00284
SIR	smallworld	0.10	0.00607	0.01531	0.0249	-0.04346	0.00171
SIR	smallworld	0.15	0.05186	0.08794	0.08288	0.05263	0.00793
SIR	smallworld	0.20	0.05635	0.15674	0.0957	0.03002	0.00901
SIS	scalefree	0.10	0.20004	0.08303	0.2042	0.05261	0.00635
SIS	scalefree	0.15	0.27653	0.09826	0.26503	0.07858	0.04579
SIS	scalefree	0.20	0.32109	0.13629	0.31345	0.05399	0.02333
SIS	smallworld	0.10	0.00761	0.02375	0.01917	-0.01027	-0.00899
SIS	smallworld	0.15	0.06719	0.15883	0.13379	0.06029	0.00663
SIS	smallworld	0.20	0.11495	0.20274	0.21729	0.02264	-0.00918

Top Ten Overlap of Unweighted Centrality Measures and Mean TKO on Flattened Network

Disease Type	Network Type	InfectionRate	Degree	Closeness	Betweenness	Eigenvector	Katz
SIR	scalefree	0.10	0.008	0.004	0.008	0.	0.
SIR	scalefree	0.15	0.008	0.016	0.004	0.	0.
SIR	scalefree	0.20	0.016	0.012	0.02	0.	0.
SIR	smallworld	0.10	0.004	0.008	0.004	0.	0.
SIR	smallworld	0.15	0.008	0.024	0.008	0.	0.
SIR	smallworld	0.20	0.016	0.008	0.016	0.	0.
SIS	scalefree	0.10	0.012	0.02	0.012	0.	0.
SIS	scalefree	0.15	0.016	0.008	0.004	0.	0.
SIS	scalefree	0.20	0.016	0.02	0.02	0.	0.
SIS	smallworld	0.10	0.004	0.004	0.016	0.	0.
SIS	smallworld	0.15	0.02	0.012	0.016	0.	0.
SIS	smallworld	0.20	0.012	0.016	0.032	0.	0.

Table B.9: The Pearson and Spearman correlations as well as the average percent of matching Top Ten agents between the mean TKO score with each of five flattened observed interaction network agent centrality scores.

Mean Delta Fraction TKO and Base Interaction Network

Pearson Correlations of Centrality Measures and Mean Delta TKO on Base Network

Disease Type	Network Type	InfectionRate	Degree	Closeness	Betweenness	Eigenvector	Katz
SIR	scalefree	0.10	0.11126	0.09799	0.08769	0.0877	0.07232
SIR	scalefree	0.15	0.14889	0.11732	0.12976	0.12374	0.11078
SIR	scalefree	0.20	0.16805	0.14606	0.13956	0.14173	0.12793
SIR	smallworld	0.10	-0.00009	0.04849	0.0285	-0.01346	-0.015
SIR	smallworld	0.15	0.07554	0.15629	0.11958	0.08021	0.0855
SIR	smallworld	0.20	0.10399	0.17922	0.14133	0.03744	0.0903
SIS	scalefree	0.10	0.1261	0.08495	0.10497	0.08869	0.06882
SIS	scalefree	0.15	0.18292	0.11692	0.14573	0.12508	0.10177
SIS	scalefree	0.20	0.24463	0.17563	0.20131	0.17664	0.14665
SIS	smallworld	0.10	0.02352	0.08397	0.02964	-0.01187	0.00138
SIS	smallworld	0.15	0.09169	0.22063	0.15242	0.04315	0.07855
SIS	smallworld	0.20	0.14818	0.2428	0.2825	0.0684	0.15092

Spearman Correlations of Centrality Measures and Mean Delta TKO on Base Network

Disease Type	Network Type	InfectionRate	Degree	Closeness	Betweenness	Eigenvector	Katz
SIR	scalefree	0.10	0.1898	0.09614	0.1907	0.08385	0.04946
SIR	scalefree	0.15	0.21847	0.12032	0.20791	0.11745	0.08826
SIR	scalefree	0.20	0.22746	0.13761	0.21875	0.1287	0.09995
SIR	smallworld	0.10	0.01018	0.03939	0.03931	0.02556	0.00609
SIR	smallworld	0.15	0.0516	0.14602	0.08077	0.0788	0.0599
SIR	smallworld	0.20	0.0699	0.21156	0.12171	0.05664	0.06476
SIS	scalefree	0.10	0.19765	0.08699	0.2017	0.07697	0.04237
SIS	scalefree	0.15	0.29223	0.10894	0.28278	0.10884	0.06348
SIS	scalefree	0.20	0.32746	0.13763	0.32083	0.12769	0.07348
SIS	smallworld	0.10	0.00888	0.05328	0.03496	-0.02485	-0.02329
SIS	smallworld	0.15	0.07207	0.24837	0.14039	0.05486	0.05916
SIS	smallworld	0.20	0.11845	0.25568	0.22694	0.07612	0.11974

Top Ten Overlap of Centrality Measures and Mean Delta TKO on Base Network

Disease Type	Network Type	InfectionRate	Degree	Closeness	Betweenness	Eigenvector	Katz
SIR	scalefree	0.10	0.016	0.012	0.012	0.	0.
SIR	scalefree	0.15	0.028	0.016	0.032	0.	0.
SIR	scalefree	0.20	0.024	0.012	0.02	0.	0.
SIR	smallworld	0.10	0.008	0.	0.	0.	0.
SIR	smallworld	0.15	0.012	0.02	0.008	0.	0.
SIR	smallworld	0.20	0.02	0.02	0.012	0.	0.
SIS	scalefree	0.10	0.012	0.024	0.016	0.	0.
SIS	scalefree	0.15	0.024	0.004	0.012	0.	0.
SIS	scalefree	0.20	0.024	0.044	0.028	0.	0.
SIS	smallworld	0.10	0.012	0.008	0.02	0.	0.
SIS	smallworld	0.15	0.012	0.02	0.032	0.	0.
SIS	smallworld	0.20	0.012	0.012	0.028	0.	0.

Table B.10: The Pearson and Spearman correlations as well as the average percent of matching Top Ten agents between the mean change in fractional TKO score with each of five base network agent centrality scores.

Mean Delta Fraction TKO and Flattened Interaction Network - Unweighted

Pearson Correlations of Unweighted Centrality Measures and Mean Delta TKO on Flattened Network

Disease Type	Network Type	InfectionRate	Degree	Closeness	Betweenness	Eigenvector	Katz
SIR	scalefree	0.10	0.11209	0.09676	0.08987	0.04575	-0.00955
SIR	scalefree	0.15	0.14898	0.11589	0.13081	0.07341	0.00383
SIR	scalefree	0.20	0.16834	0.14605	0.14042	0.05829	0.01177
SIR	smallworld	0.10	-0.00009	0.04849	0.0285	-0.03495	-0.00892
SIR	smallworld	0.15	0.07554	0.15629	0.11958	0.06106	-0.00153
SIR	smallworld	0.20	0.10399	0.17922	0.14133	-0.00044	0.00307
SIS	scalefree	0.10	0.12659	0.08274	0.10637	0.07632	0.01533
SIS	scalefree	0.15	0.18318	0.11273	0.14702	0.1164	0.02053
SIS	scalefree	0.20	0.24569	0.17587	0.20314	0.11365	0.01806
SIS	smallworld	0.10	0.02352	0.08397	0.02964	-0.00576	0.01179
SIS	smallworld	0.15	0.09169	0.22063	0.15242	0.04087	0.01275
SIS	smallworld	0.20	0.14818	0.2428	0.2825	-0.00623	0.00016

Spearman Correlations of Unweighted Centrality Measures and Mean Delta TKO on Flattened Network

Disease Type	Network Type	InfectionRate	Degree	Closeness	Betweenness	Eigenvector	Katz
SIR	scalefree	0.10	0.18978	0.09581	0.19107	0.05107	-0.02437
SIR	scalefree	0.15	0.21847	0.12017	0.21048	0.06814	0.02238
SIR	scalefree	0.20	0.22745	0.13857	0.21806	0.04818	0.00276
SIR	smallworld	0.10	0.01018	0.03939	0.03931	-0.05202	0.00721
SIR	smallworld	0.15	0.0516	0.14602	0.08077	0.09211	0.01342
SIR	smallworld	0.20	0.0699	0.21156	0.12171	0.03232	0.00737
SIS	scalefree	0.10	0.19763	0.08611	0.20295	0.05174	0.01168
SIS	scalefree	0.15	0.29223	0.10699	0.28511	0.08752	0.04597
SIS	scalefree	0.20	0.32746	0.13833	0.31985	0.05933	0.02361
SIS	smallworld	0.10	0.00888	0.05328	0.03496	-0.00993	-0.00835
SIS	smallworld	0.15	0.07207	0.24837	0.14039	0.11152	0.00933
SIS	smallworld	0.20	0.11845	0.25568	0.22694	0.01745	-0.01791

Top Ten Overlap of Unweighted Centrality Measures and Mean Delta TKO on Flattened Network

Disease Type	Network Type	InfectionRate	Degree	Closeness	Betweenness	Eigenvector	Katz
SIR	scalefree	0.10	0.016	0.02	0.012	0.	0.
SIR	scalefree	0.15	0.024	0.008	0.024	0.	0.
SIR	scalefree	0.20	0.028	0.024	0.016	0.	0.
SIR	smallworld	0.10	0.008	0.	0.	0.	0.
SIR	smallworld	0.15	0.012	0.02	0.008	0.	0.
SIR	smallworld	0.20	0.02	0.02	0.012	0.	0.
SIS	scalefree	0.10	0.012	0.016	0.012	0.	0.
SIS	scalefree	0.15	0.02	0.004	0.012	0.	0.
SIS	scalefree	0.20	0.028	0.024	0.02	0.	0.
SIS	smallworld	0.10	0.012	0.008	0.02	0.	0.
SIS	smallworld	0.15	0.012	0.02	0.032	0.	0.
SIS	smallworld	0.20	0.012	0.012	0.028	0.	0.

Table B.11: The Pearson and Spearman correlations as well as the average percent of matching Top Ten agents between the mean change in fractional TKO score with each of five flattened observed interaction network agent centrality scores.